THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIV | May, 1933 | Number 5

THE COEFFICIENT OF RANK: A TECHNIQUE FOR EVALUATING PUPILS' ABILITY

DONCASTER G. HUMM AND KATHRYN A. HUMM

Los Angeles Diagnostic Clinic

The classification of pupils for educational purposes presents a two-fold problem: First, the placement in an appropriate grade; and, second, the classification of the pupils within the grade in groups of suitable size and with a certain amount of homogeneity for classroom work.

The study here reported was undertaken in an effort to find a solution for the second problem stated above. As the study progressed, it was found that light had been shed upon some of the problems of grade placement and promotion policies as well.

This article will, therefore, first present a technique for classifying pupils within a grade into groups of approximately equal ability; and will then report, in their bearing on policies of grade placement, findings regarding educational acceleration and retardation which were brought to light by the application of this technique.

In using the expression, "equal ability groups," the authors do not wish to imply that they advocate a policy of promoting pupils in accordance with chronological age and of attempting to solve all problems of differences in ability by means of classifying the pupils in each grade into equal ability groups. No matter what policy governs the promotion of pupils from grade to grade, it is necessary in a large school system to divide the pupils in each grade into groups of suitable size for classroom teaching. Such division offers an opportunity for classification in which the ability of the pupils is the important consideration. Equal ability grouping may be practiced whether the pupils of a given grade are of the same chronological age or of the same mental age or, as is more often the case, of widely varying chronological and mental ages.

Since the study deals with questions of retardation and acceleration in school, and since these terms are used in different senses by different authors, it is necessary at the outset to define these terms as they will be used in this article. Acceleration will be used to mean placement in a grade higher than the mental age of the pupil warrants; retardation, placement in a grade lower. For example, dull children who have been promoted beyond their mental age because of advanced chronological age or large size will be said to be accelerated; while bright children who have not been advanced in grade because of their youth will be said to be retarded.

The study was made in a junior high school, comprising the seventh, eighth, and ninth grades.

It extended over a period of five years, during the first half of which pupils were tested with the Terman Group Test of Mental Ability, and during the second half with the Otis Self-administering Test of Mental Ability.* The change of tests was an administrative expedient, not connected with this study.

The data reported from the elementary schools of Los Angeles were secured from schools in which the National Intelligence Tests alone were used.*

PART I. CLASSIFICATION OF PUPILS INTO EQUAL ABILITY GROUPS

For the purpose of classifying pupils into equal ability groups, intelligence tests were given routinely at the beginning of each semester.

For the purpose of this study, records were preserved, although tabulations and distributions were made from time to time, until the number of scores totaled 2800. Final distributions were made of that number.

There are in current use two methods of classifying pupils within the grade: (1) By means of the intelligence quotient, and (2) by means of mental age.

---

* Kelley[6] has found the Otis test to correlate with the Stanford-Binet +.80; the Terman Group Test, +.75. Kelley further notes that the Otis test has a community of function with the Stanford-Binet of ninety-five per cent; and the Terman Group Test, of ninety-three per cent. He reports that retests with Otis tests gave: $r = +.87$. Otis and Terman compared gave: $r = +.84$. It seems evident that the results of the two tests may be combined without seriously affecting the findings.

The National Intelligence Tests correlate with the Stanford-Binet to the extent of +.84, with a community of function of ninety-four per cent.

At first thought, it would seem that either mental ages or intelligence quotients would afford a measure of rank in ability by which such a classification could be made. However, this was found not to be the case, since each of these methods, tried in turn, brought from teachers many complaints that the groups thus formed were not sufficiently homogeneous for teaching purposes.

Inspection of the distributions of pupils indicated that the fact that many pupils have become accelerated or retarded has introduced a confusing factor into the situation, so that, as shown by the following examples, neither intelligence quotients nor mental ages, used alone, afford a suitable measure.

Of 1240 pupils present in one school year, thirty-one had an intelligence quotient of exactly one hundred. On entering the low seventh grade, these thirty-one pupils had mental ages distributed as follows:

Mental ages (months)	165–169	160–164	155–159	150–154	145–149	140–144	135–139	130–134
No. of pupils	1	5	4	9	5	5	1	1

This represents a variation of about three years in mental age for pupils of uniform intelligence quotient.

The mean mental age for pupils entering the low seventh grade is one hundred fifty months. Of the pupils present during the year mentioned, seventy-one had entered the low seventh grade with a mental age of exactly one hundred fifty months. The intelligence quotients of these pupils were distributed as follows:

IQ's	Number of pupils	IQ's	Number of pupils	IQ's	Number of pupils
113–114	1	103–104	9	93–94	3
111–112	2	101–102	12	91–92	2
109–110	6	99–100	7	89–90	5
107–108	1	97–98	10	87–88	1
105–106	1	95–96	8	85–86	3

This table shows a variation of thirty points in intelligence quotient for pupils of uniform mental age.

These examples were found to be typical of the findings for all intelligence quotients and for all mental ages. They unmistakably indicated that in a given grade pupils of a given intelligence quotient vary in ability as indicated by mental age; and those of a given mental age, in ability as indicated by intelligence quotient.

It became apparent, therefore, that a ranking of pupils on the basis of either mental age or intelligence quotient, used alone, would be inexpedient.

Inspection of a scatter diagram brought to light two significant facts: (1) When IQ's are used as measures of rank in ability, and the pupils divided on that basis, it is the wide difference in mental age within each group which interferes with approximate homogeneity; while (2) when mental ages are used as the basis of division, it is the wide difference in IQ which interferes. In other words, in one case the variation in mental maturity within the group is the disturbing factor; in the other case, the variation in rate of mental development.

The problem, therefore, became one of finding a measure or combination of measures which would bring together in each group pupils of relatively uniform mental maturity whose mental development is proceeding at approximately the same rate.

It is, of course, to be taken for granted that complete homogeneity is neither feasible nor, probably, necessary. There will always be greater variation within the groups than there is between the extremes of adjacent groups; and there will always be variations in temperamental qualities, life aims, etc., which are not the subject of this study. It was sought in this study to secure only such homogeneity as might simplify to some extent classroom teaching problems.

The success of the method of dividing the pupils was judged on the basis of teachers' reports. Throughout the study the teachers of classes were emphatically encouraged to report to the counselor those pupils who seemed to be misfits in their classes. All pupils so reported were retested at once. If it was found that the score of the original test was incorrect, he was transferred to the appropriate group. If, on the other hand, it was found that the score of the original test was correct, his case received individual attention and study until the cause of the maladjustment was found.

Before presenting the formula which was used in classifying the pupils, it is necessary to describe the measures of ability which were used.

The Intelligence Quotient.—Since the rate of mental development is one of the factors of the problem, as stated above, and since the intelligence quotient is a measure of that rate, it was determined to retain the intelligence quotient as one of the measures to be used in classifying pupils.

It is necessary to determine the significance of the intelligence quotient, in order that the second measure to be combined with it may be exactly comparable to it.

It is customary to define the intelligence quotient as the quotient derived by dividing a subject's mental age by his chronological age. This definition is sufficiently accurate for most purposes, but for the purpose of this study it is necessary to call attention to the fact that the intelligence quotient is derived rather by dividing the subject's mental age by the mean mental age of all subjects of his chronological age.[4,8] In other words, the intelligence quotient measures rate of mental development by comparing the subject's mental age with the normal mental age of his chronological age group; or, it is a measure of relative mental maturity for people of the same chronological age.

The Ability Ratio.—A measure of relative mental maturity for pupils of the same grade in school can be found by dividing a pupil's mental age by the mean mental age of all pupils in that grade, and such a measure is, as will be shown below, comparable to the intelligence quotient. This measure was named the Ability Ratio (AbR).

Validating Comparable Expressions.—Kelley[5] states that the minimum number of conditions which must be met before two scales can be fully equated are three: (*a*) A point of the first must be known to be equal to a point of the second, (*b*) a second point of the first must be known to be equal to a second point of the second, and (*c*) the law establishing the relationship between successive points on the first must be known to be the law underlying the second.

The formulation of the two measures in question makes inevitable the meeting of the first of these conditions. Since each is an index based on a mean, each will have as its mean one hundred.

The remaining two conditions must be established by an examination of intelligence quotients and ability ratios to find their relationships.

Kelley[5] also states: "Where more than one element demands consideration at one time, some measure of their relationship is needed to assist in their simultaneous interpretation. There are two such measures available, correlation and regression, of which the latter is

statistically the more fundamental. Correlation becomes regression where deviations are equal."

Hull[3] points out the disadvantages of a multiple regression formula. He says: "The deriving of the equation is a complicated and technical procedure, and the arithmetic used is somewhat more complicated than the simple addition usually employed in test batteries."

However, if it is possible to construct one formula, especially if it be a rather simple one, it may be reduced to chart form and be made readily available. In order to accomplish this reduction it will be necessary: (*a*) To equate the deviational relationships of the intelligence quotient and the ability ratio; (*b*) to establish that these relationships are uniform for the grades in which they are to be used; (*c*) to derive the formula; and (*d*) to construct the chart.

After the chart has been constructed, it will be necessary further (*e*) to establish the relationships between the measures which indicate acceleration and retardation; and (*f*) to derive such other interpretations as may be valuable.

The comparison of the distributions of the two measures in question, undertaken below, seeks first to meet the remaining two conditions for equating two scales, and second to reduce these to a single formula on which a chart may be constructed, making available an easily applied technique for the rapid classification of pupils at the beginning of each school year.

The Distribution of the Intelligence Quotient.—Since Otis[7] points out that the scores of group intelligence tests are less accurate than those of individual intelligence tests,* it was considered necessary to make a distribution of intelligence quotients from these data, rather than to use such distributions as were already available.

The table on p. 327 shows the distribution of intelligence quotients for the period of our study.

From these facts we find:

(1) The distribution is very symmetrical, since the amount of skew is very small when compared with the amount (.50) permissible in distributions which may be treated as symmetrical distributions without essential error.

---

* It is of interest to know that care in administering group tests makes possible results of rather high accuracy. The distribution of the difference between first test IQ's and second test IQ's (one hundred twenty-eight instances) gave a mean of +1.3, a median of +.50, a mode of −1.1, Q_1 as −4.1, and Q_3 as +5.4; thus rendering a semi-interquartile range (Q) of 4.8.

TABLE I.—DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN A LOS ANGELES JUNIOR HIGH SCHOOL

IQ	Number	IQ	Number	IQ	Number
50–54	1	90–94	307	130–134	61
55–59	7	95–99	349	135–139	29
60–64	16	100–104	363	140–144	20
65–69	40	105–109	285	145–149	14
70–74	68	110–114	257	150–154	3
75–79	112	115–119	193	155–159	7
80–84	157	120–124	178	160–164	3
85–89	244	125–129	86	165–169	0

Total population, 2,800.
Mode, 99.5; mean, 101.6; median, 100.9.
Mn. Dev., 13.00; St. Dev., 16.58; Q, 11.00; D (10–90 percentile), 42.24; PE, 11.18.
Standard error of mean, .311; Probable error of mean, .21.
Coefficients of skew; Modal, .161; median, .054.
Ku (measure of kurtosis), .260 (leptokurtic).

(2) The PE, 11.18, is about 40 per cent greater than Terman's findings for intelligence quotients derived from individual intelligence tests.

(3) The measures of central tendency (Mean, 101.6; Median, 100.9; and Mode, 99.5) are generally higher than the expected findings, which are all one hundred. This may possibly be accounted for by the selectivity of higher grade work.

(4) The normal relationships of the probable error, mean deviation, and standard deviation are rather closely preserved: PE (11.18) equals .86 Mn. Dev. (13.00), as against the normal curve relationship, .845 Mn. Dev.; and Mn. Dev. (13.00) equals .784 St. Dev. (16.58), as against the normal curve relationship, .798 St. Dev.

(5) The range equals 6.3 standard deviations, less than would be expected in a normal distribution, which has a range of seven standard deviations for a population of 1085.

The Distribution of the Ability Ratio.—Since the ability ratio has for its formula:

$$\text{AbR} = \frac{\text{MA}}{\text{MGA}}*$$

*MGA = Mn · MA of the grade.

it is necessary to find accurate norms for the grade age. From the Los Angeles School Reports,[13] norms from 40,000 cases, averaging 5,000 to the grade, were as follows:

TABLE II.—MEAN MENTAL AGES BY GRADES (LOS ANGELES)

Grade	Mean MA, years	Grade	Mean MA, years
I	7.0	V	11.0
II	8.0	VI	12.0
III	9.0	VII	13.0
IV	10.0	VIII	14.0

These figures agree exactly with the estimate given by Terman in "The Intelligence of School Children."[10]

When these figures for the whole grade are interpolated for half grades, the figures given serve for the upper half grade, and figures half a year lower serve for the lower half of each grade. Thus the table would be:

TABLE III.—MEAN MENTAL AGES BY HALF GRADES (LOS ANGELES)

Grade, low	Mean MA, years	Grade, high	Mean MA, years
I	6.5	I	7.0
II	7.5	II	8.0
III	8.5	III	9.0
IV	9.5	IV	10.0
V	10.5	V	11.0
VI	11.5	VI	12.0
VII	12.5	VII	13.0
VIII	13.5	VIII	14.0
	Extrapolating		
IX	14.5	IX	15.0
X	15.5	X	16.0
XI	16.5	XI	17.0
XII	17.5	XII	18.0

Checking the above figures by the tables of mental ages prepared in the Los Angeles Elementary Schools in 1925,[1] we find:

TABLE IV.—MEAN MENTAL AGES BY HALF GRADES (LOS ANGELES 1925)

Grade, low	Mean MA, years	Difference	Grade, high	Mean MA, years	Difference
I	6.25	− .25	I	6.75	− .25
II	7.3	− .2	II	7.75	− .25
III	8.5	0.00	III	8.9	− .1
IV	9.25	− .25	IV	10.0	0.00
V	10.5	0.00	V	11.0	0.00
VI	11.6	.1	VI	12.2	.2
VII	12.8	.3	VII	13.0	0.00
VIII	13.6	.1	VIII	13.9	− .1
IX	14.4	− .1	IX	14.8	− .2
X	15.3	− .2	X	15.7	− .3
XI	16.2	− .3	XI	16.7	− .3
XII	17.0	− .5	XII	17.6	− .4
Total		−1.30	Total		−1.70
					−1.30
Grand total					−3.00
Mean difference					− .125
Median difference					− .182
Semi-interquartile range					.134

The amount by which the figures in Table IV are less or greater than those of Table III is indicated in the column marked "Difference." It can be seen that the tendency is for the ages in the second table to be a little less than the corresponding ages in the first table. The average difference is .125 years lower for the second table. If the figures for the seventh, eighth, and ninth grades (the grades with which this study is particularly concerned) are considered, the average difference between the ages in the two tables is zero; and, if the first nine grades are considered, the average difference is .055 years less for the second table. This is equal to about two-thirds of one month.

The norms of the first table may be checked further by anticipating the results to be reported a little further on. As shown there, the central tendencies of the ability ratio fall between 101.93 and 102.83, with the median at 102.53; the central tendencies of the intelligence quotient at 101. This marks a slight tendency on the part of the entire group to be superior. Since one hundred AbR and one hundred IQ represent the assumed means, these central tendencies rep-

resent a deviation of two points for the ability ratio and a deviation of one point for the intelligence quotient. Since, as is shown later in the comparison of the two distributions, the intelligence quotient and ability ratio deviate as four is to three, the ability ratio has a mean tendency to be 0.5 points higher than it should be. Thus, the norms chosen might be thought to have resulted in an error of plus 0.5 points.

That this finding reveals not an error in calculation, but rather a variation in pupils, is shown by the variation, quoted from the report by Clark,[1] of 93.8 to 117.7 IQ in mean intelligence quotient of schools. It is also shown by the variation in the school in which this study was made. For example, the median intelligence quotient of the school population of eight hundred fifty-two, taken February 2, 1925, was 103.4; the median intelligence quotient of the population of seven hundred forty-two, taken January 20, 1926, was 101.57; of the cumulative frequency of 1594, taken at the same time, 102.55. Such variations in intelligence quotient are necessarily accompanied by corresponding variations in mental age.

It seems preferable to retain the norms of the first table; first, because they are regular and seem to correspond to regular tendency; and, second, for the reason that an arbitrary point in a regular distribution may be chosen to represent a guessed mean without error,[5] especially when the data are considered in more than twelve units of frequency.

TABLE V.—DISTRIBUTION OF ABILITY RATIOS

AbR	Number	AbR	Number	AbR	Number
60–64	4	90–94	309	120–124	128
65–69	13	95–99	402	125–129	78
70–74	23	100–104	519	130–134	40
75–79	39	105–109	416	135–139	12
80–84	104	110–114	297	140–144	10
85–89	199	115–119	206	145–149	1

Total population, 2800.
Mode, 101.93; mean, 102.83; median, 102.53.
Mn. Dev., 9.7; St. Dev., 12.71; Q, 8.03; D (10–90 percentile), 32.29.
Standard error of the mean, .24; PE of the mean, .162.
Coefficients of skew: Modal, .093; median, .031.
Ku (measure of kurtosis), .248 (leptokurtic).

With these norms as a basis, the ability ratio for each pupil was calculated by the formula given above, and a distribution of all the ability ratios was made.

From these facts we find:

(1) That the distribution is very symmetrical, since the amount of the skew is considerably smaller than is allowable in distributions which may be treated as symmetrical without essential error.

(2) Q:Mn. Dev.::.828:1 (normal, .845:1), Mn. Dev.:St. Dev.:: .763:1 (normal, .789:1), Q:St. Dev.::.632:1 (normal, .674:1).

It may be noted that the slight degree of difference in the measures of dispersion as compared to those of the normal curve increases as the limits of the respective measures depart from the mean. This is due to the fact that this curve of distribution is more peaked or more leptokurtic than the normal curve.

Comparison of the Intelligence Quotient and the Ability Ratio.— Kelley[6] gives the following formula for the determination of kurtosis:

$$\text{Ku} = \frac{\text{Q}}{\text{D}}$$

in which Ku is the measure of kurtosis; Q, the quartile deviation; and D, the ten to ninety percentile range. He gives Ku = .263 as the critical measure of kurtosis, stating that if Ku is greater than .263, the distribution is platykurtic; and if it is less than .263, the distribution is leptokurtic.

Since the D of the ability ratio equals 32.29 and the Q equals 8.03, Ku for the ability ratio equals .248; and since the D of the intelligence quotient equals 42.24 and the Q equals 11.00, Ku for the intelligence quotient equals .260. Thus, the AbR distribution is leptokurtic, and the IQ distribution is also leptokurtic.

Kelley[5] also states that the assumption of linearity of regression of two distributions carries with it the necessity of equal kurtosis, if arrays are homoscedastic, that is, if the standard deviations of the *X* arrays and *Y* arrays are equal.

The problem of determining the degree of homoscedasticity in these distributions and the permissible difference in kurtosis, in other words, the linearity of the regression, may be solved practically for these two distributions to a sufficient degree of accuracy for our purposes by an examination of the relationships of the deviations as given on p. 332.

When the two distributions are compared, the following facts are established:

TABLE VI.—RELATIONSHIP OF IQ AND AbR

Q(IQ):Q(AbR)::1:.730::4:2.92
MnD(IQ):MnD(AbR)::1:.746::4:2.98
StD(IQ):StD(AbR)::1:.766::4:3.06
D(IQ):D(AbR)::1:.764:4:3.06

It is to be noted that the effect of the difference in kurtosis of the two distributions is to cause a slight departure from true linearity of regression; otherwise, all of these deviations would have the same ratio. However, if the ratio 1.00:.75, or 4:3, is chosen to express the relationship of the deviations of the IQ to those of the AbR, and weighting in that ratio is used, the result is entirely accurate except for very high and very low values. For such values, the inaccuracy will not exceed one or two points. In other words, the difference in kurtosis between the two distributions is so small that linearity may be assumed.

From these relationships we may equate an expression which will express the regression of the IQ and AbR.

The Coefficient of Rank; A Determination of Rank in Grade on the Basis of a Regression of IQ and AbR.—An equation which will express the regression of the IQ and AbR will at once express the departure from the central tendency with regard to ability as compared to the mean of the grade and as compared to the means of various chronological ages, and will thus fulfill the requirements of a double criterion. Accordingly, this equation was chosen as representing a truer rank within the grade than either of its components alone, and the measure of rank so arrived at has been called the coefficient of rank (CoR).

By assigning to the coefficient of rank formula a weighting of the ability ratio sufficient to make its deviations coincide with those of the intelligence quotient, this regression formula becomes a statement of rank in terms of intelligence quotient.

Such a formula is:

$$\text{Deviation of CoR} = \frac{\overline{\text{Deviation of IQ}} + \frac{4}{3}\overline{\text{Deviation of AbR}}}{2}$$

Since one hundred has been taken uniformly as the assumed mean (where it is not the actual mean), this formula may be evolved into:

$$100 - \text{CoR} = \frac{(100 - \text{IQ}) + \frac{4}{3}(100 - \text{AbR})}{2}$$

Solving:

$$\text{CoR} = \frac{\text{IQ}}{2} + \frac{2(\text{AbR} - 25)}{3}$$

If the weighting is correct, one would expect in such a large number of instances as are included in our data that the distribution of the coefficient of rank would nearly coincide with the distribution of the intelligence quotient. An examination of the distribution of the CoR, given below, indicates that this expectation is borne out.

TABLE VII.—DISTRIBUTION OF COEFFICIENTS OF RANK IN A LOS ANGELES JUNIOR HIGH SCHOOL

CoR	Number	CoR	Number	CoR	Number
45–49	1	85–89	221	125–129	92
50–54	3	90–94	281	130–134	65
55–59	8	95–99	359	135–139	37
60–64	17	100–104	380	140–144	27
65–69	33	105–109	331	145–149	12
70–74	56	110–114	254	150–154	5
75–79	96	115–119	186	155–159	6
80–84	151	120–124	177	160–164	2

Total population, 2800.
Mode, 100.8; mean, 102.3; median, 101.8.
Mn. Dev., 12.70; St. Dev., 16.55; Q, 10.85; D (10–90 percentile), 41.84.
Standard error of mean, .312; Probable error of mean, .211.
Coefficients of skew: Modal, .118; median, .030.
Ku (measure of kurtosis), .269 (platykurtic).

To even a casual inspection there is strong similarity between the distribution of the intelligence quotient and that of the coefficient of rank.

The table on p. 335 of agreement brings this similarity out in detail.

This is essential agreement within very narrow limits. The central tendencies are all close to one per cent of agreement; the measures of distribution, within one unit. The amount of skew in both is inconsiderable. In the light of the agreement of these measures, we have a justification of the processes which led to these results.

A Chart for the Computation of Coefficients of Rank.—To facilitate the calculations necessary to the determination of this measure of rank, a calculation chart was computed and drawn. It was possible to

THE CALCULATION OF RANK
Lon B. Klemm — 1928
ABILITY AGES (MA OR Ed.A) SCALED BY GRADES
DIRECTIONS: Find the Ability Age (MA or Ed.A) on the scale designated below by the pupil's grade, and the C.A. on the scale designated on the left by the same grade. From the former project down and from the latter to the right. The intersection indicates the value of the IQ and the Coefficient of Rank (C+R)
9th Grade High
9th Grade Low
8th Grade High
8th Grade Low
7th Grade High
7th Grade Low
— ABILITY —
— RATIOS —
(increase)
REGRESSION TENDENCY
Chronological Ages by Grades
9th Grade High
9th Grade Low
8th Grade High
8th Grade Low
7th Grade High
7th Grade Low
CHR = CA / GA
IQ = MA / CA
AHR = MA / GA
CHRT = (Chronological Ratio)
AHR = Ability Ratio
GA = Median Age of Grade (smoothed distribution)
C+R = Coefficient of Rank

TABLE VIII.—COMPARISON OF THE DISTRIBUTIONS OF IQ'S AND CoR'S

Measure	IQ	CoR	Difference (basis, IQ)
Mode	99.5	100.8	1.3
Mean	101.6	102.3	.7
Median	100.9	101.8	.9
Mn. Dev	13.02	12.7	− .3
St. Dev	16.58	16.55	− .03
PE (mean)	.210	.211	.001
Q	11.00	10.85	− .15
D	42.24	41.84	− .40
Kurtosis	.260	.269	.009
Skew: Modal	.161	.118	− .043
Skew: Median	.054	.039	− .015

design this chart so that intelligence quotients, ability ratios, and coefficients of rank could be computed simultaneously.

In order to construct a chart which would be applicable to all the school grades, it was necessary to use another expression, which was named the chronological ratio (ChR). The chronological ratio has for its formula:

$$\text{ChR} = \frac{\text{CA}}{\text{ChGA}}$$

in which CA signifies the chronological age of the pupil, and ChGA the mean chronological age of pupils of his grade (both ages being used as of the beginning of the school term).

TABLE IX.—TABLE OF MEAN CHRONOLOGICAL AGES BY GRADES (SMOOTHED)

Grade, low	MnCA		Grade, high	MnCA	
	Years	Months		Years	Months
I	6	6	I	7	0
II	7	6	II	8	0
III	8	6	III	9	0
IV	9	6	IV	10	0
V	10	6	V	11	0
VI	11	6	VI	12	0
VII	12	6	VII	13	0
VIII	13	6	VIII	14	0
IX	14	6	IX	15	0

The mean chronological ages for each grade were found from cumulative frequencies compiled by the author from distributions made by Herbert Popenoe[9] and Willis W. Clark[1,2] in the Los Angeles Schools over a period of several years. The chronological ages so found are our proposed norms for this formula shown in Table IX on p. 335.

As in the case of the ability ratio, but to an even greater extent, the number of cases was so large that the probable error of the mean became inconsiderable, and the mean could be used as a constant.

TABLE X.—DISTRIBUTION OF CHRONOLOGICAL RATIOS

ChR	Number	ChR	Number
75–79	1	105–109	244
80–84	9	110–114	103
85–89	48	115–119	68
90–94	127	120–124	27
95–99	307	125–129	13
100–104	292	130–134	1

Total population, 1240.
Mode, 102.21; mean, 102.45; median, 102.19.
Mean Dev., 6.48; St. Dev., 8.05; Q, 5.30; D (10–90 percentile), 21.67.
Standard error of the mean, .231; PE (mean), .149.
Coefficient of skew: Modal, .07; median, .07.
Ku (coefficient of kurtosis), .244 (Leptokurtic).

Although the proposed norms do not exactly fit the findings of any given year, we may use them with safety, for, as in the case of the ability ratio, the accuracy of the results is not materially influenced by the use of a guessed mean rather than the true mean in handling the statistics, especially when the mean has been derived from a distribution of relatively fine grouping.[5]

TABLE XI.—COMPARISON OF DISTRIBUTION OF ChR WITH NORMAL DISTRIBUTION

Q: MnDev.::.818:1 (normal, .845:1)
MnDev.: St. Dev.::.805:1 (normal, .789:1)
Q: St. Dev.::.658:1 (normal, .674:1)

In order to justify the use of the chronological ratio with the ability ratio in the construction of our chart, it is necessary to present the same information concerning it which has already been presented for the ability ratio.

TABLE XII.—COMPARISON OF DISTRIBUTION OF ChR WITH DISTRIBUTION OF AbR

$Q_a : Q_e :: 1 : .657 :: 3 : 2.00$
$MnD_a : MnD_e :: 1 : .668 :: 3 : 2.01$
$StD_a : StD_e :: 1 : .632 :: 3 : 1.90$
$D_a : D_e :: 1 : .674 :: 3 : 2.02$

From these figures, we find that the distribution is symmetrical, since the skew is small. We also find a rather close agreement between the relationships of the measures of dispersion of the ChR as compared with normal distribution, from which we may infer a high degree of homoscedasticity and homoclisy.

It will be observed that the chart for the calculation of rank, and other measures, is scaled horizontally by chronological ratios and vertically by ability ratios. For convenience, the scaling has not followed convention; for the ability ratios decrease from left to right; and the chronological ratios decrease from bottom to top. Above the ability ratios the equivalent mental ages for each grade of the junior high school are scaled; similarly, to the left of the chronological ratios, the chronological ages for each grade of the junior high school are scaled.

Broken oblique lines represent the charting of equal intelligence quotients; and solid oblique lines, which are slightly curved, represent the charting of equal coefficients of rank.

The projection vertically from the pupil's ability ratio to the intersection with the horizontal projection of his chronological ratio will mark the point at which his intelligence quotient and coefficient of rank may be read.

Intelligence quotients and coefficients of rank are scaled for even numbers only. The odd numbers must be read as halfway between two even numbers.

If a pupil's ability ratio equals ninety and his chronological ratio also equals ninety, the intersection will fall on the one hundred IQ line and will be read one hundred IQ; but the intersection also falls between the one hundred two CoR line and the one hundred four CoR line, and should be read one hundred three CoR.

The scaling of equivalent mental ages for each grade opposite the ability ratios and of equivalent chronological ages for each grade opposite the chronological ages makes it possible to obtain intelligence quotients and coefficients of rank from these ages without computing the respective ratios. Thus, a low seventh grade pupil with a chronological age of eleven years three months and a mental age of thirteen years

nine months may be evaluated by following the lines passing through the corresponding AbR (one hundred ten) and the corresponding ChR (ninety) to their intersection, to be read one hundred twenty-two IQ and one hundred eighteen CoR.

The Use of the Regression Formula for the Classification of Pupils in Equal Ability Groups.—The application of the formula presented here is much simpler than its statistical validation. It is necessary only to give an intelligence test to all the pupils, to determine their mental ages, and to verify their chronological ages. The intelligence quotient and the coefficient of rank for each pupil may then be ascertained from the chart, as described heretofore.

The pupils in each grade may be divided into equal ability groups by ranking them in accordance with their coefficients of rank, and dividing the whole range into groups of such size as may be suitable for the classrooms, etc., available.

In the school in which this technique was developed (the average daily attendance of which was approximately eight hundred), this entire task was completed on the first day of the semester. Pupils previously in attendance had already been assigned to their new classes at the end of the previous semester. Group intelligence tests were administered to all new pupils in the morning by the counselor. The completed tests were scored at once by a corps of four or five teachers experienced in that work, who determined the mental ages and verified the chronological ages. The counselor then computed the ability ratio and coefficient of rank for each new pupil and assigned him to the appropriate class.

As new pupils entered during the semester, they were immediately tested and assigned to classes.

The value of the method was found in the results secured by its use. Whereas, when mental age alone was used for classification, pupils referred back to the counselor for reclassification were frequently found to be misplaced because of discrepancies in intelligence quotient between themselves and other members of their groups; and when intelligence quotient alone was used, pupils referred back were frequently found to be misplaced because of discrepancies in mental age between themselves and other members of their groups; when the method of classification by means of the coefficient of rank was used, pupils were still sometimes referred back to the counselor for reassignment, but in these cases the cause for maladjustment was found in every case to be either an error in the score of the intelligence test

itself, inadequate preparation for the work of the grade, ill health, or some other of the numerous causes of individual maladjustment which may occur. The classes, as a whole, have been found to be more satisfactorily arranged in "equal ability" groups.

PART II. APPLICATION OF THE COEFFICIENT OF RANK TO THE STUDY OF GRADE PLACEMENT

As soon as mental tests began to be used in the study of educational problems, important findings were brought to light with regard to the relative progress through school of pupils of various degrees of brightness.

Terman, in "The Intelligence of School Children,"[10] presents the following summary of some of these findings:

"1. The lower the IQ, the greater the degree of retardation* . . .

"2. The IQ groups above one hundred show a greater degree of acceleration* the brighter the group. It will be noted, however, that the acceleration of the bright group is not quite as great as the retardation of the dull . . .

"3. If the mental age of a given child in one of the retarded groups is computed, it will usually be found that the child is less retarded than he ought to be. When the mental age of a child in one of the bright groups is computed, it will be found that the child is less accelerated than he ought to be.

"4. The typical child of sixty or sixty-five IQ tends to remain in the first grade until the age of ten or eleven years, and not to reach the fifth grade until the age of fourteen or fifteen years. By this time he has a mental level of only about nine years and is not able to do the school work satisfactorily above the third or fourth grade.

"5. The typical child of seventy-five to seventy-nine IQ reaches the fifth grade by the age of thirteen years, and if he remains in school is likely to be found in the eighth grade by the age of sixteen or seventeen. Nearly always, however, his grade location is higher than the mental age would warrant.

"6. Children of eighty to eighty-four IQ usually remain two years in the first grade, and complete the eighth grade, if they complete it at all, one or two years behind schedule time.

---

* Note that the terms, acceleration and retardation, are used by Terman with reference to chronological age. Elsewhere in this article they are used with reference to mental age.

"7. On the other hand, children of one hundred twenty to one hundred twenty-nine IQ are usually found either one or two grades accelerated. Nearly all of this gain, however, is made in the first year or two of school life. After the first year, they are held to the one-grade-one-year pace of average children. Even so, the central tendency is for them to complete the eighth grade at the age of thirteen.

"8. The situation is slightly but not proportionately better for the IQ group of one hundred thirty to one hundred thirty-nine. Children of one hundred forty to one hundred seventy IQ, however, are likely to become three or four years accelerated and to reach the eighth grade by the age of eleven or twelve years. Wherever children of the higher IQ groups are located, their work always presents a striking contrast with that of children of the sixty, seventy, or eighty IQ class who are several years their seniors."

Terman's statement that the retardation of dull children and the acceleration of bright children (with reference to chronological age) is in both cases less than their mental age would warrant is confirmed by the data collected for our study.

A comparison of the distribution of the ChR (page 336) with that of the AbR (page 330) shows that the dispersion of the latter is uniformly about fifty per cent greater, the deviation of the ChR being about two-thirds that of the AbR. In other words, there is fifty per cent greater departure from the mean of the grade with reference to a mental age standard than with reference to a chronological age standard.

The data which were accumulated for the statistical validation of the coefficient of rank, described in Part I, revealed to us findings concerning acceleration and retardation of students with reference to their mental age, which confirm and supplement previous findings in this field. These findings are presented in Part II.

The Relation between Intelligence and Grade Placement.—On the chart for the calculation of rank, a line which intersects the graphs representing equivalent deviational tendencies of the IQ and AbR marks the central tendency of placement for respective IQ's and AbR's; that is to say, the line will cross each IQ graph at the point of central tendency for AbR's of pupils of that IQ and will cross each AbR graph at the point of central tendency for IQ's of pupils of that AbR. Statistically speaking, this is the line of best fit. Since this line indicates a definite tendency for pupils of a given IQ to reach a given grade at a

certain mental age, or for pupils of a given mental age who have reached a given grade to have a certain IQ, it reveals the rate of promotion for each such group and may be appropriately designated the line of promotion tendency.

In a school in which the mean IQ is one hundred and the mean AbR is one hundred, the line of promotion tendency will pass through the intersection of those two graphs. In the school in which this study was made, the mean IQ was one hundred one and the mean AbR one hundred two. The line of promotion tendency for that school, therefore, as shown on the chart, page 334, passes through the point of intersection of one hundred one IQ and one hundred two AbR.*

To make more readily accessible the information revealed by the line of promotion tendency, we have prepared the table on p. 342, showing the chronological and mental age at which the pupil of each IQ group reaches each grade in the junior high school, and also showing the grades for which such ages are normal.

From this table it may be seen that the average pupil of the grade of intelligence classified as "genius" is four to five years retarded according to mental age while in junior high chosol; the very superior pupil, two to four years retarded; the superior pupil, one to two years retarded; the normal pupil, less than one year retarded or accelerated; the dull normal pupil, one to two years accelerated; the pupil of borderline intelligence, two to three years accelerated; and the mentally deficient pupil, three or more years accelerated.

To indicate what proportion of pupils are thus affected, a table has been prepared showing the percentage of pupils to be found in each of these classifications.

The figures in Table XIV indicate that in the junior high school studied, a little less than half of the pupils were in the grade corresponding to their mental age, a little less than one-fourth were in a grade beyond that corresponding to their mental age, and a little less than one-third were in a grade below that corresponding to their mental age.

* It is to be noticed that, as drawn on the chart, the line of promotion tendency is a curve. This is not due to curvilinear regression, but to the construction of the chart. The IQ graphs are oblique lines converging toward zero mental age and zero chronological age, while the AbR graphs are parallel vertical lines. The regression expressed with such an arrangement of lines results as in the projection of a line from a curved surface. When the IQ graphs and AbR graphs are so arranged as to be perpendicular to each other, the line of promotion tendency becomes a straight line.

TABLE XIII.—CHRONOLOGICAL AND MENTAL AGES AT WHICH PUPILS OF EACH IQ GROUP REACH EACH JUNIOR HIGH SCHOOL GRADE, WITH GRADES FOR WHICH THESE AGES ARE NORMAL

IQ	CA	Grade for which CA is normal	MA	Grade for which MA is normal	CA	Grade for which CA is normal	MA	Grade for which MA is normal
	Low Seventh				High Seventh			
60–69	14–6	Low 9	9–5	High 3	15–2	High 9	9–9	Low 4
70–79	13–8	Low 8	10–4	High 4	14–5	High 8	10–9	Low 5
80–89	13–4	High 7	11–3	High 5	13–10	Low 8	11–8	Low 6
90–99	12–11	Low 7	12–3	High 6	13–5	High 7	12–9	Low 7
100–109	12–7	Low 7	13–1	High 7	13–1	High 7	13–8	Low 8
110–119	12–4	High 6	14–1	High 8	12–8	Low 7	14–8	Low 9
120–129	12–1	High 6	15–2	High 9	12–7	Low 7	15–9	Low 10
130–139	11–11	Low 6	16–1	High 10	12–5	High 6	16–9	Low 11
140–149	11–9	Low 6	17–1	High 11	12–3	High 6	17–5	High 11
	Low Eighth				High Eighth			
60–69	15–8	Low 10	10–1	High 4	16–3	High 10	10–6	Low 5
70–79	14–11	Low 9	11–2	High 5	15–6	Low 10	11–7	Low 6
80–89	14–5	Low 8	12–2	High 6	14–11	Low 9	12–7	Low 7
90–99	14–0	High 8	13–3	Low 8	14–6	Low 9	13–9	Low 8
100–109	13–7	Low 8	14–2	High 8	14–1	High 8	14–8	Low 9
110–119	13–4	High 7	15–3	High 9	13–10	Low 8	15–10	Low 10
120–129	13–1	High 7	16–4	High 10	13–7	Low 8	16–11	Low 11
130–139	12–11	Low 7	17–4	High 11	13–4	High 7	18–0	High 12
140–149	12–9	Low 7	19–1	College	13–2	High 7	19–9	College
	Low Ninth				High Ninth			
60–69	16–11	Low 11	10–10	Low 5	17–5	High 11	11–3	High 5
70–79	16–1	High 10	12–0	High 6	16–7	Low 11	12–5	High 6
80–89	15–6	Low 10	13–0	High 7	16–0	High 10	13–6	Low 8
90–99	15–0	High 9	14–2	High 8	15–6	Low 10	14–8	Low 9
100–109	14–7	Low 9	15–3	High 9	15–1	High 9	15–9	Low 10
110–119	14–4	High 8	16–5	High 10	14–10	Low 9	16–11	Low 11
120–129	14–0	High 8	17–6	Low 12	14–6	Low 9	18–1	High 12
130–139	13–10	Low 8	18–9	College	14–4	High 8	19–4	College
140–149	13–7	Low 8	20–	College	14–1	High 8	20–	College

Although these figures indicate a very serious amount of educational misplacement in the school under consideration, they would not be worth the consideration of educators generally if they do not repre-

TABLE XIV.—PERCENTAGE OF PUPILS IN EACH INTELLIGENCE CLASSIFICATION WITH APPROXIMATE MISPLACEMENT IN GRADE ACCORDING TO MENTAL AGE

IQ	Classification	Per cent of pupils	Acceleration or retardation by mental age, years	
69 and less	Moron	2.3	3 to 4	Accelerated
70–79	Borderline	6.4	2 to 3	
80–89	Dull normal	14.3	1 to 2	
Total below normal		23.0		
90–99	Normal	23.4	0 to 1	
100–109	Normal	23.2	0 to 1	
Total normal		46.6		
110–119	Superior	16.1	1 to 2	Retarded
120–129	Very superior	9.4	2 to 3	
130–139	Very superior	3.2	3 to 4	
140 and over	"Genius"	1.7	4 to 5	
Total above normal		30.4		

sent a general tendency. That they do represent more or less accurately such a tendency may be inferred from the fact that the pupils entering this school were not the product of a single educational system, but had attended schools in every part of the United States and some English-speaking foreign countries. Pupils from non-English-speaking countries were not included in the study. This variety of educational background was due to the circumstance that the school was located in a part of Los Angeles given over mainly to apartment houses, where newcomers to the city lived for a few months while selecting the places for their permanent homes.

To determine to what extent the tendency in this school corresponded with the tendency in Los Angeles as a whole, data of the Los Angeles elementary schools were secured through the courtesy of Willis W. Clark, and were converted into ability ratios to facilitate comparison.

The following table presents the mean deviation of the ability ratio in the several grades, together with the mean mean deviation. The schools from which these data were secured were selected as representative, having their mean intelligence quotients in the neighborhood of one hundred. Schools with mean intelligence quotients as high as one hundred five were used, and were compensated for by schools of equal deviation from one hundred in the opposite direction. With this exception, the selection of the schools was at random.

TABLE XV.—MEAN DEVIATION OF ABILITY RATIO IN ELEMENTARY GRADES OF LOS ANGELES SCHOOLS

Grade	Mn. Dev. AbR	No. of pupils	Dev. weighted
B1	10.4	911	93.6
A1	10.0	811	80.0
B2	10.1	876	90.9
A2	9.3	823	74.4
B3	9.4	771	75.2
A3	8.8	719	61.6
B4	9.4	768	75.2
A4	8.7	694	60.9
B5	9.4	744	65.8
A5	9.3	594	55.8
B6	9.6	614	57.6
A6	10.8	476	54.0
B7	9.2	252	27.6
A7	7.8	242	15.6
B8	9.1	154	18.2
A8	9.8	95	9.8
Total		9544	916.2

Mean Mean Dev. AbR = 916.2 ÷ 95 (total weights) = 9.6

It may be seen from this tabulation and calculation that the mean mean deviation of the ability ratio in the Los Angeles elementary schools varies but .1 point of an AbR from the mean deviation found in the junior high school under consideration, since the latter is 9.7 points and the former 9.6. None of the elementary grades varies in its mean AbR from the mean mean of the whole distribution by more than 1.7 AbR points. The average difference is .52 AbR point.

It seems quite clear, therefore, that the retardation and acceleration tendencies of the school investigated are similar to those of the Los Angeles schools in general.

Attention is called to the fact that there is no uniformity of departure from the mean deviation of 9.6 AbR which can be ascribed to lowness or highness of grade.

The constancy of the AbR is similar to the constancy of the intelligence quotient, but is not to be interpreted in the same manner; for the intelligence quotient of an individual does not seem to be capable of being greatly changed, except under certain abnormal conditions, while the ability ratio may be changed at once by promotion to a higher grade or demotion to a lower. The constancy of the intelligence quotient is due to the rate of maturation, while the constancy of the ability ratio must be ascribed to the inertia of the grade system of promotion.

There is a relation of factors. The tendency of pupils of higher intelligence quotient to forge ahead of the average in mental development, and of the pupils of lower intelligence quotient to lag behind, under systems which take imperfect cognizance of these tendencies and which make little provision for differentiating them in terms of accelerated or retarded progress, cause a damming up within the grades of retarded and accelerated pupils in a fairly constant ratio to the intelligence quotient.

It is not the purpose of this paper to outline in any detail methods of handling the situation which has been revealed. The policies which are advocated by various educators may be roughly classified into two groups, those which attempt to provide differentiated instruction in each grade in order to meet the needs of pupils of wide ranges of ability, and those which attempt to keep educational progress approximately in step with mental growth.

With regard to the former policy, the question arises whether the subject-matter prescribed by the curriculum can be taught to all the pupils of a given grade. For example, it is essential in the seventh grade that the dull pupils receive instruction in the third, fourth and fifth grade levels, while the bright pupils are ready for work in the senior high school level. It is doubtful whether the subject-matter which is appropriate for the average pupil in the seventh grade can be learned at all, no matter how simply taught, by the dull pupils in that grade, or can be made sufficiently difficult to command the attention of the very bright pupils.

It would seem, therefore, that the continuation of the policy of promoting pupils at the rate of a grade a year without taking cogni-

zance of their rate of mental development would necessitate differentiated curricula within the grades.

A very common objection to the promotion of pupils at the rate of their mental development is the fact that such a policy would result in some pupils reaching the higher grades while still very young. It is feared that this would result in serious social maladjustment. Whether social maladjustment would result from educating pupils of widely different chronological ages together to a greater extent than it may arise from educating together pupils of widely different mental ages is a question that must be decided by further research. Attention may be called here, however, to the fact that a general policy of promoting pupils in accordance with mental age would result in there being enough young bright pupils in the higher grades and enough older dull pupils in the lower grades to permit of their being segregated into special groups, if that were found to be desirable.

It may be observed by referring to Table XIII that neither policy is in full effect, since there is a wide range of chronological ages and a wider range of mental ages in each grade.

SUMMARY

This study was undertaken for the purpose of finding a method of classifying pupils of a given grade into groups of suitable size and of sufficient homogeneity for classroom teaching. The method which was developed provided a means of making a school survey which revealed the amount of retardation and acceleration of pupils classified by intelligence quotient, and which brought to light important questions regarding policies of grade placement and rate of promotion.

Part I.—The classification of pupils in a given grade by intelligence quotient for purposes of classroom teaching was found to be unsatisfactory because pupils of a given intelligence quotient in a given grade have widely varying mental ages, while classification by mental age was found to be unsatisfactory because pupils of a given mental age in a given grade have widely varying intelligence quotients.

In order to take into consideration both relative mental maturity and relative rate of mental development, a formula was used by which rank in grade was determined on the basis of the regression of intelligence quotient and ability ratio (the ability ratio is derived by dividing the pupil's mental age by the mean mental age of pupils in his grade in school). The formula is:

$$\text{coefficient of rank} = \frac{\text{IQ}}{2} + \frac{2(\text{AbR} - 25)}{3}$$

By means of this formula, a chart was constructed by which the coefficients of rank of pupils in a grade could be readily ascertained, making possible the classification of entering pupils with a minimal expenditure of time and effort.

Part II.—The data which were accumulated for the statistical validation of the formula given above revealed the following findings concerning acceleration and retardation of pupils in the school studied:

1. The deviation of the ability ratio is about fifty per cent greater than the deviation of the chronological age ratio (derived by dividing the pupil's chronological age by the mean chronological age of pupils in his grade in school). This indicates that, while pupils in a given grade are not homogeneous with respect to chronological age, they are even less homogeneous with respect to ability as indicated by mental age.

2. On the chart for the calculation of rank, the line of best fit may be constructed. This line indicates a definite tendency for pupils of a given intelligence quotient to reach a given grade at a certain mental age, and therefore reveals the average rate of progress through school of pupils classified according to intelligence. It was found in the junior high school studied that bright pupils were retarded about one year for every ten points by which their intelligence quotients exceeded the mean, while dull pupils were accelerated about one year for every ten points by which their intelligence quotients fell below the mean.

3. It was found, further, that only 46.6 per cent of pupils were in a grade corresponding to their mental ages; twenty-three per cent of the pupils were from one to four years beyond the grades corresponding to their mental ages; and 30.4 per cent of pupils were from one to five years below the grades corresponding to their mental ages.

4. When computed for elementary grades in representative schools in the same city, the ability ratio showed a relatively constant deviational tendency throughout the grades (mean deviation = 9.6), indicating that the findings presented above apply also to these elementary schools.

Although the study here reported was made in a single junior high school and the findings verified by comparison with data secured from other schools in the same city, the fact of acceleration of dull pupils and retardation of bright pupils is well known, and it is not unlikely that studies of other schools and school systems would reveal that the

extent of retardation and acceleration found in this school is fairly typical.

If further researches reveal this to be the case, one or the other of the two possible policies must be followed: Either every grade must be so organized as to offer instruction appropriate to pupils ranging in mental age from five years below to four years above the mean of the grade, and the term "grade" must refer only to the number of years a pupil has been in school; or pupils must be placed in grade according to mental age and must be promoted at the rate of their mental development.

BIBLIOGRAPHY

1. Clark, Willis W.: Educational Status of Los Angeles Elementary Schools. *Educational Research Bulletin, Los Angeles City Schools*, Vol. IV, No. 2, January 15, 1925.
2. ——: Age-grade Standards for Los Angeles City Schools. *Los Angeles School Journal*, Vol. VI, No. 13, December 4, 1922.
3. Hull, Clark L.: "Aptitude Testing." World Book Company, Yonkers-on-Hudson, New York, 1928.
4. Humm, Doncaster G.: A Chart for the Better Evaluation of Intelligence Quotients. *Journal of Applied Psychology*, Vol. XVI, No. 1, February, 1932, pp. 91–96.
5. Kelley, Truman L.: "Statistical Method." The Macmillan Company, New York, 1922.
6. ——: "Interpretation of Educational Measurements." World Book Company, Yonkers-on-Hudson, New York, 1927.
7. Otis, Arthur S.: "Self-administering Tests of Mental Ability, Manual of Directions and Key." World Book Company, Yonkers-on-Hudson, New York, 1922.
8. ——: "Statistical Method in Educational Measurement." World Book Company, Yonkers-on-Hudson, New York, 1926.
9. Popenoe, Herbert: "Distribution of Mental Ages in the Elementary Grades." Report to Supt. of Schools, Los Angeles, Calif., Dec. 5, 1924.
10. Terman, Lewis M.: "The Intelligence of School Children." Houghton Mifflin Co., Boston, 1919.
11. ——: "Group Test of Mental Ability." World Book Co., Yonkers-on-Hudson, New York, 1924.

RELATIVE VALUES OF OBJECTIVE AND ESSAY TYPE EXAMINATIONS IN GENERAL PSYCHOLOGY

A. R. GILLILAND AND L. E. MISBACH

Northwestern University

As Cason[1] has recently pointed out, experimental evidence concerning the relative values of objective and essay type examinations in college courses is not yet conclusive. The following study is a comparison for one college subject (psychology) of the two types of examination in terms of correlations between psychology grades and intelligence test scores and between psychology grades and semester grade averages. It is assumed that an important consideration in determining the value of a type of examination is the extent to which it differentiates students in respect to general academic ability, as indicated by intelligence test scores and semester grade averages.

A further question which the data presented below bear upon is whether or not essay and objective tests measure the same quality or kind of achievement. Semester grade averages are based very largely on essay tests, objective tests being used in very few college courses. If essay tests measure some peculiar quality of achievement not measured by objective tests, there should be an advantage in favor of grades based on essay tests in correlation with semester grade averages. It is also of interest to find whether or not a study of correlation coefficients indicates any peculiar common quality in objective tests and the intelligence test.

METHOD OF STUDY

The data for this study were obtained from the records of fourteen classes in general psychology, selected from classes of several semesters in the Northwestern University College of Liberal Arts. The classes were selected with a view to obtaining in respect to size of classes, identity of instructors, quality of instruction, and testing methods used the most accurate comparison possible from the records available between results obtained with essay and objective type examinations. In seven classes grades were based on essay tests only, and in seven other classes grades were based on objective tests only.

The classes selected were taught by four experienced instructors of proved ability. Each instructor chose on the basis of his own prefer-

[1] Cason, H.: The Essay Examination and the New Type Test. *School and Society*, Vol. XXXIV, 1931, pp. 413–418.

ence the type of examinations used in his classes, having no foreknowledge of this study. One instructor, *A*, used only essay examinations. Two instructors, *B* and *C*, used essay examinations in one class each and objective tests in other classes. Instructor *D* used only objective tests.

For all Liberal Arts students in the fourteen classes chosen, a total of nine hundred forty students, the following data were obtained from the college records: Semester grades in psychology, percentile rankings on the American Council of Education Psychological Examination, and the index values of the average grade of each student in all courses for the same semester in which psychology was taken.

Percentile rankings on the intelligence test were transmuted into sigma scores, only the sigma scores being used for correlations.

Index values of semester grade averages as obtained from the college records were recomputed with psychology excluded so as to give the average grade for the semester in all courses other than psychology. Only these recomputed semester grade averages were used for correlation with psychology grades.[1]

TABLE I.—CLASSES FROM WHICH DATA WERE OBTAINED

Classes given essay examinations					Classes given objective examinations				
Instructor	Class	Year	Total number students	Intelligence test score and semester average available for	Instructor	Class	Year	Total number students	Intelligence test score and semester average available for
A	I	1928–1929 2nd. sem.	102	73	*D*	VIII	1928–1929 1st. sem.	53	40
A	II	1929–1930 1st. sem.	85	50	*D*	IX	1929–1930 1st. sem.	90	58
A	III	1930–1931 1st. sem.	88	53	*D*	X	1930–1931 1st. sem.	127	89
A	IV	1931–1932 1st. sem.	85	51	*B*	XI	1928–1929 1st. sem.	83	49
A	V	1930 summer	68	55 (30)*	*B*	XII	1929–1930 1st. sem.	68	42
B	VI	1924–1925 1st. sem.	81	79	*B*	XIII	1930–1931 1st. sem.	131	90
C	VII	1929–1930 1st. sem.	135	106	*C*	XIV	1930–1931 1st. sem.	137	104

* Semester grade averages were obtainable for fifty-five students in Class V, but intelligence test scores were obtainable for only thirty.

[1] Because the College in 1930 adopted a new system of computing index values from grades, making the weighting of grades in relation to each other different

For eleven of the fourteen classes used, instructors' grade-books giving a record of scores on individual examinations were also available.

General features of the classes from which data were obtained are shown in Table I.

TYPES OF EXAMINATIONS AND GRADING METHODS

Essay examinations were of the traditional type, being heavily weighted with discussion questions. Classes I, II, III, IV, and V were given from two to three one-hour tests during the semester, each test consisting of five to ten questions, and a final examination approximately two hours in length, consisting of twenty to thirty questions. Papers for Class I were graded by one assistant, numerical scores being assigned each paper. Papers for Classes II, III, and IV were graded by an assistant who used a rough rating method, assigning letter grades. This assistant did not record grades and had no contact with the classes. Papers for Class V were graded by the instructor, who used a rating method, assigning only letter grades. Grades for Class VI are based on approximately four one-hour essay examinations. Papers for this class were graded by the instructor, numerical grades being given. Class VII was given two or three one-hour essay tests and a two-hour final examination which was in part objective. Papers for this class were graded by several assistants, numerical scores being assigned.

The objective examinations for the most part were composed of about one-half true-false items and about one-half recall and recognition items. The score on each paper was merely the number of items answered correctly, no correction for chance being made. Most students marked all true-false and recognition items, guessing on those items which were not known. In Classes IX, X, XI, XII, XIII, and XIV, taught by instructors, *B*, *C*, and *D*, the same examinations were used in classes occurring in each year. New examinations were made

---

from the relative weighting in the older system, it was necessary, in order to combine for correlation purposes classes of several years, that index values of semester grade averages be transmuted into some common unit system. This was done by throwing the recomputed (psychology excluded) index values of the older system, for all classes of this study occurring prior to 1930–1931, into a distribution table and finding the mean and SD of this distribution, then transmuting all recomputed index values of the older system into sigma scores. The same transmutation was then made for index values of the new system. The recomputed semester grade averages were, thus, for all correlations with psychology grades, expressed as sigma scores.

for each year. In these classes a short examination, consisting of about thirty items and requiring from fifteen to twenty minutes of the students' time was usually given each week, and a final examination about 2 hours in length was given.

RELIABILITY OF TESTS[1]

As the records available did not include files of examination papers, direct measurement of the reliability of the examinations was not possible. However, for all classes except II, VII, and XIV the instructors' records of grades given on examinations were available. For five classes given essay tests, the consistency of grades for each class was found by averaging grades given on two one-hour examinations during the semester and finding the correlation between these averages and grades given on the two-hour final examination. For six classes given weekly objective tests, scores on about six such tests were averaged as being equivalent to two one-hour examinations. Correlations found are shown in Table II.

The reliability of these tests obtained by a split-half method would presumably be higher than the correlation obtained as above between examinations taken at different times in the semester. However, comparisons of the correlation coefficients for different classes should indicate the relative reliability of the tests. It is possible that in two classes, Classes V and VI, where papers were graded by the instructor, a halo effect due to personal contact with the students may give an artificially high reliability to grades. This is not the case in Classes I, III, and IV, since the assistants who graded papers for those classes did not record grades and had no contact with the classes, having personal acquaintance with not more than four or five students in each class.

It appears that the objective tests attain significantly greater reliability than the essay tests, though there is considerable overlapping. The very low correlation coefficient found for Class I may serve to indicate that the reliability of an essay type examination depends very much on the qualifications and carefulness of the grader

[1] Correlation between tests taken at different times in the semester is here called reliability, in conformity with the usage of Paterson (*School and Society*, Vol. XXIV, 1926, pp. 246–248) and others. It must be pointed out, however, that this is actually only an indirect and relative measure of test reliability as measured by more exact methods.

TABLE II.—CORRELATION BETWEEN AVERAGE OF SCORES ON TWO ONE-HOUR TESTS AND SCORES ON TWO-HOUR FINAL TEST

Instructor	Class	Type of test	Number of students	Method of grading papers	r*
A	I	Essay	102	Papers given numerical scores. Grading by one assistant	+.16 ± .065
A	III	Essay	88	Papers rated and given letter grades. Grading by one assistant	+.55 ± .05
A	IV	Essay	85	Papers rated and given letter grades. Grading by one assistant	+.56 ± .052
A	V	Essay	68	Papers rated and given letter grades. Grading by instructor	+.68 ± .044
B	VI	Essay	81	Papers given numerical scores. Grading by instructor	+.54 ± .053
Correlation between average of scores on tests equivalent to two one-hour tests and scores on two-hour final test					
D	VIII	Objective	53	Score on each test, no. items correct. Scoring by assistants	+.61 ± .058
D	IX	Objective	90	Same as Class VIII	+.76 ± .03
D	X	Objective	127	Same as Class VIII	+.75 ± .026
B	XI	Objective	83	Same as Class VIII	+.56 ± .051
B	XII	Objective	68	Same as Class VIII	+.74 ± .037
B	XIII	Objective	131	Same as Class VIII	+.57 ± .04

* Computed by product-moment method.

and on the methods of grading used. Brinkley,[1] Sims,[2] and others have found that essay tests may attain fairly high reliability if rating methods of grading are used.

EQUILIZATION OF EXAMINATIONS IN RESPECT TO LENGTH AND DISTRIBUTION OF GRADES

In order to equate the two types of tests so that comparisons might have a more exact basis, grades were taken for each class, as far as possible, from tests which required a standard amount of time on the part of the student. This does not, of course, equate the tests in

[1] Brinkley, S. G.: Values of New-type Examinations in the High School. New York: Teachers College, Columbia University, Contributions to Education, 1924, p. 161.

[2] Sims, Verner M.: The Objectivity, Reliability and Validity of an Essay Examination Graded by Rating. *Journal of Educational Research*, Vol. XXIV, 1931, pp. 216–223.

respect to time required for constructing and scoring. In classes given essay tests, the smallest number of hours used for testing was about four, while no class used more than five hours. In the classes given objective tests, about six hours per semester were used in taking examinations. Complete records were available for all classes given objective tests except Class XIV, and in order to equate the length of the objective tests with that of the essay tests, new total scores were computed for Classes VIII, IX, X, XI, XII, and XIII, omitting scores made on about half of the weekly objective tests given during the semester. This made the total of objective tests in these classes roughly equivalent to two one-hour examinations of about one hundred items each and a two-hour final of over two hundred items. New total scores thus obtained were distributed into letter grades, the distribution being made as nearly as possible to correspond to the distribution of grades for the combined classes of the same year in which essay examinations were used.

This re-distribution of objective test scores into irregular letter-grade intervals disregards the fact that one advantage of the objective test is that the wide range of scores obtained makes normal distribution of grades comparatively easier. In the present case, however, it was found that, although in most classes re-distribution of objective test scores as described above materially altered correlations of grades with the criteria, the average effect for all classes was insignificant. On the whole it seemed best for the purposes of this study to assume that grades on essay examinations can be distributed according to the instructor's desire, and to attempt to equate grades based on the two types of examinations in every possible manner, so that wider distribution of scores or greater length of examinations would not give an advantage to either type of test.

For all classes but II, VII, and XIV, the correlations between scores on a single test, the two-hour final examination, and intelligence test scores and semester grade averages were found. Since objective tests are much more comprehensive per unit of time than essay tests,[1] it was thought that important comparisons might be made between results obtained with two-hour objective tests and four-hour essay tests, and that the effects of shortening the tests might give some clue to the value of large sampling, especially in the essay test.

---

[1] Talbot, E. O. and G. M. Ruch: Minor Studies on Objective Examination Methods. *Journal of Educational Research*, Vol. XX, 1929, pp. 199–206.

RESULTS AND DISCUSSION OF RESULTS

Correlations between grades and intelligence test scores and semester grade averages are shown in Table III. Correlations are shown for grades based on both four-hour and two-hour testings. For convenience, reliability coefficients given in Table II are reproduced in Table III.

CORRELATION BETWEEN INTELLIGENCE TEST SCORES AND GRADES BASED ON FOUR-HOUR EXAMINATIONS

The correlation coefficients shown in Table III indicate that, in general, objective tests in general psychology differentiate students according to scores made on the intelligence test more effectively than do essay tests which are equal in length to the objective tests (equal in respect to time required for students to take the test). Differences in correlation coefficients are small but consistent.[1] Since semester grade averages have not been found to correlate much above +.50 with intelligence test scores, correlation between grades in one course and intelligence test scores will ordinarily fall considerably below +.50. The relatively small correlations between psychology grades and intelligence test scores preclude the appearance of statistically significant differences between correlations except for very large numbers of cases.

For the combination of all classes given four-hour essay tests the obtained correlation between grades and intelligence test scores is $+.20 \pm .031$; for the combination of all classes given objective tests the corresponding r is $+.31 \pm .028$. The difference of .11 between the two coefficients is 2.8 times the PE of the difference. The chances that the difference is a real difference are ninety-seven in one hundred.

It is, of course, possible that for these combinations of classes results are affected by differences in teaching methods of the instructors. Study of correlation coefficients obtained for individual classes indicates, however, that the apparent advantage of objective tests appears consistently regardless of differences between instructors. In classes given essay tests, correlations found between grades and intelligence test scores, ranging from $-.12 \pm .121$ to $+.33 \pm .07$, show much greater variability than is shown in corresponding correla-

[1] Similar results for high school courses are reported by Davis, R. A. and L. W. Johnson. The Influence of Intelligence upon Performance in New-type and Old-type Examinations. *Educational Method*, Vol. X, 1931, pp. 488–493.

TABLE III.—CORRELATIONS[1] BETWEEN PSYCHOLOGY GRADES AND INTELLIGENCE TEST SCORES AND SEMESTER GRADE AVERAGES[2]

Instructor	Class	Type of test	Number cases	Reliability (from Table II)	r between grades on four-hr. test and intelligence test scores	r between grades on two-hr. test and intelligence test scores	r between grades on four-hr. test and semester grade averages	r between grades on two-hr. test and semester grade averages
A	I	Essay	73	+.16 ± .065	+.33 ± .070	+.15 ± .077	+.48 ± .061	+.37 ± .068
A	II	Essay	50	.....	+.32 ± .086	.....	+.46 ± .075	
A	III	Essay	53	+.55 ± .05	+.26 ± .086	+.12 ± .091	+.57 ± .063	+.51 ± .069
A	IV	Essay	51	+.56 ± .052	+.09 ± .094	+.16 ± .092	+.65 ± .055	+.63 ± .057
A	V	Essay	(30)[3] (55)	+.68 ± .044	−.12 ± .121	−.22 ± .117	+.42 ± .075	+.40 ± .076
B	VI	Essay	79	+.54 ± .053	+.23 ± .072	+.25 ± .071	+.43 ± .062	+.35 ± .067
C	VII	Essay	106	.....	+.17 ± .064	.....	+.56 ± .045	
A[4], *B*, *C*	I, II, III, IV, V, VI, VII	Essay	(442)[3] (467)	....	+.20 ± .031	....	+.49 ± .024	
A[4], *B*	I, III, IV[5], V, VI	Essay	(286)[3] (311)	....	+.20 ± .038	+.15 ± .039	+.51 ± .027	+.42 ± .032
D	VIII	Objective	40	+.61 ± .058	+.36 ± .093	+.35 ± .094	+.48 ± .082	+.57 ± .072
D	IX	Objective	58	+.76 ± .03	+.33 ± .079	+.30 ± .081	+.66 ± .05	+.56 ± .061
D	X	Objective	89	+.75 ± .026	+.39 ± .061	+.36 ± .062	+.67 ± .039	+.62 ± .044
B	XI	Objective	49	+.56 ± .051	+.31 ± .087	+.41 ± .08	+.42 ± .079	+.54 ± .068
B	XII	Objective	42	+.74 ± .037	+.39 ± .088	+.42 ± .086	+.63 ± .063	+.69 ± .055
B	XIII	Objective	90	+.57 ± .04	+.30 ± .065	+.30 ± .065	+.58 ± .047	+.48 ± .055
C	XIV	Objective	104	.....	+.23 ± .063	.....	+.64 ± .039	
D[4], *C*, *B*	VIII, IX, X, XI, XII, XIII, XIV	Objective	472	.....	+.31 ± .028	.....	+.60 ± .02	
D[4], *B*	VIII, IX[6] X, XI, XII, XIII	Objective	368	.....	+.33 ± .031	+.35 ± .031	+.59 ± .023	+.56 ± .024

[1] All correlations computed by product-moment method.

[2] Average of grades in all courses other than psychology for same semester in which psychology was taken.

[3] Semester grade averages were available for fifty-five students in Class V, but intelligence test scores were available for only thirty.

[4] It should be noted that correlations of combined classes are not identica with the values which would be obtained by averaging r's.

[5] Classes for which two-hour essay test scores were available.

[6] Classes for which two-hour objective test scores were available.

tion coefficients for classes given objective tests, the range in the latter case being from $+.23 \pm .063$ to $+.39 \pm .061$.

It is of interest, also, that the lowest and highest correlations between grades and intelligence test scores for classes given essay tests occur in classes taught by one instructor, *A*. It appears, therefore, that even when classes are taught by one instructor and papers for several classes are graded by the same assistant using a constant grading method, as was the case in Classes II, III, and IV, the correlations of grades with intelligence test scores show little consistency.

Comparisons between classes taught by the same instructor where different types of tests were used also show consistently higher correlation coefficients where objective tests were used. The correlation between grades and intelligence test scores for Class VI, taught by instructor *B*, is $+.23 \pm .072$ while the lowest coefficient for any class taught by *B* when objective tests were used is $+.30 \pm .065$. Similarly, the correlation between grades and intelligence test scores for Class VII is $+.17 \pm .064$, while the corresponding correlation coefficient for Class XIV, taught by the same instructor, *C*, using objective tests, is $+.23 \pm .063$.

CORRELATION BETWEEN SEMESTER GRADE AVERAGES AND GRADES BASED ON FOUR-HOUR EXAMINATIONS

Correlations found between grades and semester grade averages are significantly higher when grades are based on objective tests than they are when grades are based on essay tests.

For the combined classes in which essay tests were used, the correlation between grades based of four-hour tests and semester grade average is $+.49 \pm .024$. The corresponding correlation for combined classes in which objective tests were used is $+.60 \pm .02$. The difference of .11 between these coefficients is more than four times the PE of the difference.

That the higher correlation of objective tests with semester grade averages exists regardless of differences between instructors is indicated by obtained correlations between grades and semester grade averages for classes taught by instructors *B* and *C*. For classes taught by *B*, in two of the three classes given objective tests, Classes XII and XIII, the obtained correlations between grades and semester grade average are respectively .20 and .15 above the similar correlation coefficient for Class VI, where essay tests were used, while in one class given objective tests, XI, the obtained correlation between grades and

semester grade average is .01 below the coefficient for Class VI. For classes taught by *C*, the obtained correlation between grades and semester grade averages is $+.56 \pm .045$ where essay tests were used and $+.64 \pm .039$ where objective tests were used.

The range of obtained correlations between grades and semester grade averages for classes given essay tests is, however, approximately identical with the range of coefficients for classes given objective tests.

CORRELATION BETWEEN CRITERIA AND GRADES ON TWO-HOUR TESTS

The higher correlation of objective test grades with criteria is more marked for two-hour than for four-hour tests. For the combination of all classes for which records of a two-hour essay test were available, the correlation between grades on the two-hour test and intelligence test scores is $+.15 \pm .039$. For a similar combination of classes given objective tests, the correlation between grades on a two-hour test and intelligence test scores is $+.35 \pm .031$. The difference between these coefficients is approximately four times the PE of the difference. Similarly, the difference between the two-hour objective tests and the two-hour essay tests in respect to correlation coefficients found between grades and semester grade averages is 3.5 times the PE of the difference.

In order to make comparisons possible between two-hour and four-hour tests for combinations of classes for which two-hour test records were available, the correlations between four-hour test grades and criteria were also found. From a comparison of these correlation coefficients it appears that for essay tests the correlations of grades with criteria tend to be lower and more variable with two-hour than with four-hour tests. On the other hand, for objective tests correlations between grades and criteria are approximately the same, both in respect to the correlation coefficients of combined classes and in respect to consistency of coefficients for separate classes, whether grades are based on four-hour or two-hour objective tests. It is of interest to note that for two-hour essay tests the correlation between grades and intelligence test scores is in no case, either for single classes or for combined classes, as large as four times its PE. On the other hand, with similar N's, in only two classes given objective tests are correlation coefficients less than four times their respective PE's.

It follows, of course, from the fact that grades on two-hour objective tests correlate as highly with criteria as grades on four-hour

objective tests, that the two-hour objective test is, in respect to such correlation, superior to the four-hour essay test.

THE SIGNIFICANCE OF RELIABILITY COEFFICIENTS

The lower tendency and wider variability of reliability coefficients (see Table II) for essay tests as compared with objective tests and the evidence, summarized above, to the effect that two-hour essay tests do not compare favorably, in respect to correlation between grades and criteria, with four-hour essay tests suggest that the most important difference between essay and objective tests is the relative unreliability of the former.

It is probable that a two-hour objective test is sufficiently comprehensive so that lengthening the test results in little gain in reliability.

It is of interest to note that the only negative correlations, $-.12 \pm .121$ and $-.22 \pm .117$, between grades and intelligence test scores occur in Class V, where the reliability coefficient is highest for all classes given essay tests. The correlations found between grades in this class and semester grade averages are also relatively small. This may indicate that the reliability of tests in this class is due to a halo effect, the grader being familiar with the students. Such results do not lend support to the argument, often advanced, that the most valuable essay test is the one given and scored by the instructor.

SIGNIFICANCE OF HIGHER CORRELATION BETWEEN OBJECTIVE TEST GRADES AND CRITERIA

The fact that grades based on the objective type examination are found to correlate more highly than grades on essay examinations with both intelligence test scores and semester grade averages, the advantage of the objective test being of the same order of magnitude[1] in both cases, seems, in respect to college psychology, to support the argument of Paterson[2] and others that the two types of test measure the same thing.

---

[1] Actually the advantage of the objective test is slightly more significant in correlation with semester grade averages than in correlation with intelligence test scores. This appears if the PE of the difference between correlations is considered, or if relative predictive values, coefficients of alienation, are considered.

[2] Paterson, D. G.: Do New and Old Type Examinations Measure Different Mental Functions? *School and Society*, Vol. XXIV, 1926, pp. 246–248.

As pointed out above, grades in most college courses are based on essay tests. It would therefore be expected, if essay tests measure some peculiar ability to achieve not measured by objective tests, that psychology grades on essay tests might correlate with semester grade averages more highly than would grades based on objective tests, or at least that the advantage of the objective test might be more evident in correlation of grades with intelligence test scores than in correlation with semester grade averages.

The data presented above indicate, however, that until some improvement, probably an increase in reliability, can be made in essay tests, convincing evidence for any peculiar value in essay tests will be lacking.

One argument which might be proposed to account for the fact that grades based on objective tests correlate more highly with intelligence test scores is that the objective test is somewhat similar in external form to the intelligence test; but the fact that the objective test grades correlate more highly than essay test grades with semester grade averages as well invalidates this argument.

It may conceivably be argued in defense of the essay test that ability to achieve in psychology is so little related to ability to succeed in an intelligence test or in general college work that correlation of grades with either of these criteria is not to be desired. Insofar as the evidence of this study bears upon this matter, it does not support such a defense of the essay type examination since, as shown in Table III, correlation between grades and criteria tends to increase as the number or length of essay tests increases. It is probable that scores on eight hours of essay testing would correlate more highly with the criteria than do scores on four-hour essay tests, just as there is an advantage in this respect of the four-hour over the two-hour essay test. Hence the more comprehensive or "better" the essay test becomes, the more it would be likely to differentiate students according to general academic ability as indicated by the criteria used here.

The lower and more variable correlations between grades based on essay tests and the criteria mean, therefore, that the four-hour essay test is a less accurate measure of achievement in psychology than is a four-hour or even a two-hour objective test. One factor in the failure of better students to receive consistently higher grades than poorer students on essay test is that grades are inaccurately determined.

CONCLUSIONS

Reliabilities of grades indicate that objective tests in general psychology attain higher and more consistent reliability than essay tests. The reliability of essay tests fluctuates very widely with different graders and different grading methods.

Grades based on objective tests in general psychology are found to correlate consistently more highly than grades based on equally long essay tests with intelligence test scores and with semester grade averages.

Psychology grades based on two-hour objective tests are found to correlate as highly as grades based on four-hour objective tests with intelligence test scores and semester grade averages. On the other hand, grades based on two-hour essay tests do not correlate as highly with the criteria mentioned above as do grades based on four-hour essay tests.

As far as correlation of grades with intelligence and with general scholastic ability may be taken as a measure of the value and meaning of tests, the data presented indicate that in the college course in general psychology, essay and objective tests measure the same quality of achievement, but that the objective test measures achievement more effectively than does the essay test.

AN ATTEMPT TO STUDY THE EFFECT OF SCIENTIFIC TRAINING UPON PREJUDICE AND ILLOGICALITY OF THOUGHT

JAMES HUNTLEY SINCLAIR AND RUTH SHERMAN TOLMAN

Occidental College

PROBLEM

"It is one of the functions of the college to promote open-mindedness—a quality sorely needed today in a period of intolerance that is worse than any other in history."[1]

This quotation from President Hopkins of Dartmouth College is representative of the recognition on the part of educators of the responsibility of the college to develop tolerance and fairmindedness in its population. In fact, it is probably fairly generally agreed in academic circles that colleges not only should but do attempt to meet this need. Partly, at least, in the assumption that a knowledge of science will further this end, twelve units of required science in lower division with laboratory is a universal hurdle for candidates seeking an arts degree. The aim to develop at least some understanding and training in the use of the scientific method is one of the reasons for the existence of the requirement. Can it be said that scientific training tends to develop suspended judgment, tolerance, and habits of reasonable thinking not only in the field of the specific science involved but also in other more general fields, such as ethics, politics, economics, etc.? To put it in another way, do individuals trained more or less intensively in science and mathematics tend to be more logical, more reasonable, less affected by presuppositions and prejudices when considering common economic, political, social and religious issues than do others of equal educational experience, but trained generally in liberal arts?

The writers, in a preliminary investigation, have attempted to discover evidence of the effects of training and proficiency in science and mathematics upon judgments on social, political and economic issues in the case of small groups of selected students. Two groups of college students from neighboring institutions were selected: One from Occidental College, a liberal arts college, and one from the California Institute of Technology, a school of science and engineering. Both

[1] Watson, G. B.: "The Measurement of Fair-mindedness." Teachers College, Columbia University, 1925.

of these draw their undergraduate student body approximately from the same geographical regions and economic levels. The former institution has about six hundred seventy-five students. Approximately one-half of these are men, who have had by the time of graduation from a minimum of twelve units up to a major in science (about thirty units). The California Institute has an undergraduate population of about five hundred thirty men, all enrolled in courses in science or engineering.

The subjects included in the Occidental group consisted of fifty-three freshmen and fifty-three seniors, all men. Of the California Institute men, forty-nine were freshmen and forty-four seniors. In the case of the freshmen, specialization in the different branches of science and engineering had not begun; in the case of the seniors, fourteen were specializing in electrical, ten in civil, and twenty in mechanical engineering.

Students entering Occidental College are selected on the basis of fifteen recommended units of high school work. For entrance to the Institute an examination is required which emphasizes preparation in mathematics and science.

The test of reasonableness used for this preliminary investigation was one prepared by Goodwin Watson, called a Survey of Public Opinion on Some Religious and Economic Issues.[1]

The Watson test consists of six parts or "Forms." Form A, the first part, is a cross-out test with fifty-one stimulus words, such as "Bolshevist," "cigarettes," "fundamentalist," "prohibition," and the like, chosen because of their emotional associations and controversial connections. The instructions are to cross out any word which is disagreeable or annoying or distasteful, and the underlying assumption is that the tendency to find a large number of these words "annoying" is indicative of prejudice in the fields indicated.

Form B, the second part, is a "Degree of Truth Test," in which the subject is requested to indicate his opinion on each of fifty-three statements by encircling one of five numbers ranging from +2 to −2. A typical example of the statements is: "The churches are more sympathetic with capital than with labor." Plus 2 represents a position of confidence so extreme as to mean that "the statement is utterly and unqualifiedly true, so that no one with a fair understanding of the subject could sincerely and honestly believe it false." The next three numbers represent a more moderate position, while −2 indicates a

[1] Bureau of Publications, Teachers College, Columbia University.

belief of extreme confidence in the falsity of the statement. The assumption is that an individual who holds such extreme views on doubtful issues as to pronounce incompetent or insincere those who differ from him tends to be prejudiced in these matters.

The third section or Form C is the "Inference Test" which was expected to be especially significant in this study for the reason that logical inference constitutes an essential part of scientific training. Its nature is indicated by the directions which precede it. They are as follows:

"Mere facts may mean different things to different people. It is often important to know just what people think certain facts mean. In the following pages you will find several statements of fact, and after each, some conclusions which people would draw from them.

"Put a check in front of each conclusion that you believe is fairly based upon the fact as given here. Do not assume anything else than the evidence given in the statement here, with all its terms understood. You are not to consider whether the conclusions are right or true in themselves, but only *whether they are rightly inferred* from the facts given in the statement."

Ten different situations are presented with from four to eight conclusions for each situation of varying degrees of connection with the facts and offering opportunities for choices dictated by bias rather than logic. The subject is protected from the necessity of making his choice from statements none of which may seem to him absolutely correct inferences by the inclusion in each group of the statement "None of these conclusions can fairly be drawn." The Inference Test seemed to the writers not only the most significant test of the Watson series, but even more significant than the whole test for analyzing out those particular features of scientific training which might differentiate the students of the two institutions.

The fourth part, Form D, is a "Moral Judgment Test." The subject is asked to pass judgment on similar methods of behavior directed toward different goals. The assumption is that if an individual approves, for example, the action of a church in distributing leaflets advertising its services but condemns a similar advertising program on the part of a group of socialists, that person is influenced in his judgment by some degree of prejudice leading to this inconsistency.

The fifth section, Form E, the "Arguments Test," presents six arguments on each of twelve issues, such as "Is Socialism desirable in the United States today?" Arguments which are both weak and

strong are presented on both sides of the question, and each is preceded by a choice of "strong" or "weak." The designation of *all* the arguments on one side as strong, of all on the other as weak, is interpreted as evidence of some prejudice on that issue.

The sixth part, Form F, the "Generalization Test," contains thirty-two statements, such as "Religious beliefs are unscientific," preceded by "all," "most," "many," "few," "no," one of which is to be encircled by the subject. According to Watson, the test rests upon what may be termed the "one swallow makes a summer" fallacy, and registers the tendency to attribute to all the members of a group characteristics true only of a part of the group. This failure to admit exceptions is assumed to be dictated by bias.

The test is scored throughout in such a way that the individual is penalized for all responses supposedly dictated by prejudice rather than by tolerant and fair-minded analysis. One is penalized one point for each word crossed out, three points for each extreme statement of degree of truth, four points for each unjustified conclusion or inference, three or five points for each inconsistency in moral judgment, four points for rating all the arguments on one side of a question weak, all on the other side strong, three points for each "all" or "no" generalization. The highest possible prejudice score is represented by a total of four hundred eighty-nine points. The score used generally is in terms of percentage, representing actual score in points divided by four hundred eighty-nine. It is clear that a low score, as the test is devised, is a superior one, representing in the total score greater freedom from prejudice and in the Inference Test fewer penalizations for incorrect inferences.

Whereas it seems to us unquestionable that a low score on the Inference Test is a desideratum and that education should tend to lower this score, we feel less certain that this would be true of the test as a whole. It seems entirely possible that many intelligent and wise men might have strong convictions on some of the questions to which they were supposed to give tentative answers, and it does not seem certain that it is one of the functions of education to eradicate these tendencies.

The norms reported by Watson are not especially useful for comparative purposes in connection with this study, since they are in large part for groups not strictly comparable to ours. Further, the numbers represented by his groups are too small to give them much significance. The group most nearly comparable for which Watson reports norms is

one of thirty-nine college graduates. The mean prejudice score for this group (gross score in percentage) is 26.6 with a standard deviation of 12.7. For the Inference Test alone he records for another group (one hundred thirty-three cases of mixed classification) a mean score of 27.1 and a standard deviation of 20.5. Reliability coefficients for the total test of .96 and for the Inference Test of .85 are reported by Watson.

RESULTS AND DISCUSSION

The results of the present study are given in the accompanying Tables I to V.

TABLE I.—NORMS FOR PERCENTAGE PREJUDICE SCORE (TOTAL SCORE)

Group	Number	Range	Mean	P.E.$_{mean}$	Median	S.D.
Occidental freshmen.....	53	8–46	29.15	.79	29.66	8.58
Occidental seniors.....	53	7–48	25.04	.80	23.57	8.60
C.I.T. freshmen.....	49	7–45	23.29	.82	22.50	8.54
C.I.T. seniors.....	44	8–44	23.77	.97	22.80	9.54

TABLE II.—GROUP DIFFERENCES: PERCENTAGE PREJUDICE SCORE

Groups compared	Differences in mean scores	P.E.$_{diff.}$	$\frac{\text{Diff.}}{\text{P.E.}_{diff.}}$
Occidental freshmen and seniors.....	4.11	1.12	3.67
C.I.T. freshmen and seniors.....	.48	1.27	.38
Occidental freshmen and C.I.T. freshmen.....	5.86	1.14	5.14
Occidental seniors and C.I.T. seniors.....	1.27	1.26	1.01

An examination of Tables I and II shows that the mean of Occidental freshmen in percentage prejudice score (29.15 per cent) exceeds the mean of Occidental seniors (25.04 per cent) by 4.11 per cent. That is, the percentage of prejudice score of the former is distinctly and significantly greater than that of the latter group. This would seem to indicate that three years of liberal arts work was somewhat effective in decreasing the tendency to reach conclusions based upon prejudice. A comparison of the mean of the California Institute freshmen with that of the seniors of the same institution reveals no significant difference. In both cases they were superior to the corresponding Occidental group, to the extent of 5.86 per cent when the comparison was

made between freshmen, but only of 1.27 per cent when the comparison was made between seniors. The large and significant difference between the freshmen suggests a difference in the type of students entering the two institutions, while the much smaller difference between the senior groups might indicate that the training of the Occidental students was bringing them nearer to the position of the Institute groups in respect to prejudice.

TABLE III.—NORMS FOR INFERENCE TEST SCORE

Group	Number	Range	Mean	P.E.$_{mean}$	Median	S.D.
Occidental freshmen.....	53	0–84	30.3	1.56	30.25	16.9
Occidental seniors.....	53	0–76	24.2	1.68	21.75	18.10
C.I.T. freshmen.....	49	0–56	19.4	1.32	17.50	13.65
C.I.T. seniors.....	44	0–60	15.6	1.22	14.28	12.05

TABLE IV.—GROUP DIFFERENCES: INFERENCE TEST SCORE

Groups compared	Differences in mean scores	P.E.$_{diff.}$	$\frac{\text{Diff.}}{\text{P.E.}_{diff.}}$
Occidental freshmen and seniors.....	6.1	2.29	2.66
C.I.T. freshmen and seniors.....	3.8	1.80	2.11
Occidental freshmen and C.I.T. freshmen.....	10.9	2.04	5.34
Occidental seniors and C.I.T. seniors.....	8.6	2.08	4.13

A study of Tables III and IV giving Inference Test scores for the four groups and their differences arranged in pairs reveals apparently significant differences. In the first comparison in Table IV, that of Occidental freshmen and seniors, the mean of the latter is 6.1 points lower (*i.e.*, better) than that of the former, a difference which is 2.66 times the probable error. This indicates statistically that the chances are 97 in 100 that a real difference in this direction exists. In the case of the comparison between California Institute freshmen and seniors, the corresponding difference in means (3.8) is 2.11 times its probable error and consequently somewhat less significant. But here again a comparison between the groups of the two institutions shows by far the largest and most significant differences. In the case of the freshmen we find a difference of 10.9 points in favor of the Institute group, which is 5.34 times its probable error; and in the case of the seniors, a difference in the same direction of 8.6 points, which is 4.13

times its probable error. These differences, which are relatively large and statistically significant, lend further support to the belief that the factor of original selection is the most significant one. It will be observed also that the means of both the Institute groups (19.4 and 15.6) on the Inference Test score are significantly lower than that given by Watson of 27.1 for one hundred thirty-three cases of mixed classification. We could assert tentatively on the basis of these results that students who elect science or engineering as their career show a tendency toward superiority in inference-making ability.

TABLE V.—CORRELATIONS

Correlations between	Group	Number	r	P.E.$_r$
Thorndike Intelligence Examination Score and Inference Test Score....	Occidental freshmen	53	−.49	.07
Thorndike Intelligence Examination Score and Percentage Prejudice....	Occidental freshmen	53	−.26	.09
Units of science and mathematics studied in college and Inference Test Score.....	Occidental seniors	53	−.08	.09

Table V presents correlations found between Thorndike Intelligence Test scores and the Inference scores, between the Thorndike score and the Percentage Prejudice score for the Occidental freshmen, and between the units of science and mathematics completed by the Occidental seniors and their Inference Test scores. The correlations were limited to the Occidental groups because no Thorndike scores were available for the Institute groups, and variation in the amounts of mathematics and science studied was significant only in the case of the Occidental seniors.

An evaluation of the coefficients found indicates that the correlation between Thorndike scores and Inference scores for the Occidental freshmen ($-.49 \pm .07$) has significance and suggests that intelligence favors superiority in inference-making ability. The coefficient of correlation between the Thorndike score and the Percentage Prejudice for the same group is too small ($-.26$) to be significant, especially since there were only fifty-three subjects in the group; and finally the correlation of $-.08$ between the amount of science and mathematics completed in terms of units by the Occidental seniors and the Inference score, while it is entirely without statistical significance, is at least

interesting as indicating no greater ability in drawing correct inferences as a result of the more prolonged study of mathematics and science,—that is, to the extent that the Inference Test is a measure of that ability.

SUMMARY AND TENTATIVE CONCLUSIONS

In the tendency to be prejudiced in drawing conclusions on such questions as are presented in the Goodwin Watson Survey of Public Opinion, this study shows differences between selected groups of students in a liberal arts college (Occidental College) and other selected groups in a technological institution (California Institute of Technology) representing rigid selection in preparatory mathematics and science and subsequent intensive training in these subjects. These differences, as measured by the Percentage Prejudice score of the Watson test, were large between the freshman groups of the two institutions (5.86 per cent), significantly smaller between the senior groups of the two institutions (1.27 per cent). This result would suggest that the training received by the Occidental students was bringing them closer to the position of the students at the Institute, and that the rating of the Institute students would not be much changed by further training. This suggestion is further supported by comparison of the Institute freshman and senior averages (23.29 and 23.77 respectively) showing no gain and even an insignificant loss for the seniors, and also by a comparison of the freshman and senior averages of the liberal arts college (29.15 and 25.04 respectively), showing an improvement of 4.11 per cent by the seniors in their tendencies to make unprejudiced judgments as measured by the test. No great significance can be attributed to these figures, however, as the groups were not large and the Goodwin Watson test is not sufficiently standardized to make it entirely trustworthy.

Since we believed the Inference Test to be both the most significant one in the Watson series and also the most useful in isolating methods of thinking closely associated with scientific training, special attention was directed to the scores on this test. The results showed large and significant differences *between* the groups of the two institutions—a difference between means of 10.9 points in the case of the freshman groups, a difference of 8.6 points in the case of the senior groups, both differences more than four times the probable errors. A comparison of freshman and senior groups *within* each of the two institutions shows

for the Occidental seniors a slightly greater gain (6.1 points) than for the Institute seniors (3.8 points).

These findings, together with a coefficient of correlation of zero determined between number of units of science and mathematics studied and Inference scores for the fifty-three Occidental seniors, do not lend support to the assumption that mathematics and scientific training are the potent factors in differentiating the groups of the two institutions in their inference-making abilities.

In such an inconclusive study any interpretations must be very tentative. Perhaps our own conclusions can be best stated negatively: That we have found here no evidence for the "transfer" of logicality or reasonable habits of thought from fields of specific scientific training to fields involving questions of economic, ethical, and social judgments as they are presented in the Goodwin Watson test. The results of the study show that there is a correlation between the choice of the scientific vocation and the ability to draw correct inferences. There is, however, no indication that this correlation is to be ascribed to the scientific training itself, and we think it not unreasonable that the pursuit of science should attract a group on the whole superior in this respect.

THE COMPARATIVE VALUE FOR JUNIOR HIGH SCHOOL USE OF FIVE GROUP MENTAL TESTS

AUSTIN H. TURNEY AND MARY FEE

University of Kansas

The number of studies having some bearing upon the value of mental tests for school uses is too large to attempt to summarize. Every study showing the correlation between mental ages, IQ's, or raw scores and school achievement as measured by marks, teachers' judgments, or other criteria, serves to answer in part the question with which we are concerned. There are, however, relatively few studies such as that of Sangren[1] which seek to determine the comparative value of a number of tests for a given level of the school as judged by a number of criteria applied to each of the tests compared. This paper is concerned with the comparative value for junior high school use of five group mental tests.

Among the reasons frequently set forth in justification of the junior high school an important place is given to individual differences and methods for adapting the school to such differences. This being true the question as to what test or tests it is best to use in the junior high school is a relatively practical and important one. It is not contended that the following data settle such a question, but they may, perhaps, be of some service to those interested in the use of mental tests in the junior high school. We wish also to point out that such evidence of validity as is presented is in no sense to be regarded as other than tentative. The report of Ascher[2] showing the results of the application of Spearman's criterion in the construction of mental tests is an example of a type of research which may well serve to question the common methods of "validating" mental tests.

METHOD

In May 1931, five widely used mental tests were administered to 190 pupils who would be entering the seventh grade of the junior high school the following September. The tests were applied by teachers who were experienced in test administration and scored and checked

[1] Sangren, Paul V.: Comparative Validity of Primary Intelligence Tests. *Journal of Applied Psychology*, Vol. 13, 1929, pp. 394–412.

[2] Ascher, E. J.: The Predictive Value of Mental Tests that Satisfy Spearman's Tetrad Criterion. *Journal of Applied Psychology*, Vol. 13, April, 1929, pp. 152–158.

by persons with special training in mental measurements. The tests used were:

1. The Terman Group Test of Mental Ability, Examination Form *B*.
2. Haggerty Intelligence Examination, Delta II.
3. McCall Multi-mental Scale.
4. National Intelligence Test, Scale *A*, Form I.
5. Otis Self-administering Tests of Mental Ability, Intermediate Examination, Form *B*.

In addition to the data obtained from these five tests this study includes data from the Stanford Revision of the Binet scale for one hundred twenty-five of the pupils. These data were on record at the time the group tests were administered. These individual examinations had been given at different times during the previous school life of the one hundred twenty-five children involved. From these, mental ages were calculated as were the mental ages for the group tests for September 1931. During the first week in September the Stanford Achievement Tests were applied to this group of pupils by the teachers in charge and these scores were made available for use in this study. Other data used were the semester letter marks for each pupil assigned by the teachers of reading, of arithmetic, and of social science. A composite rating for each pupil was also used which consisted of the average of the judgments of seven teachers. These judgments were a ranking from one to six, and corresponded to the ability section or group in which the teacher would place the pupil; a rating of "one" indicated that the pupil should be located in the group expected to do the most successful work, and a rating of "six" indicated that the pupil should be located in the group which could be expected to do the least successful work, relative to the entire grade. The judgments were made by each teacher quite independently and only by teachers who had had class experience with the pupils rated for at least three months.

RESULTS

Table I shows the distribution of the mental ages obtained by one hundred ninety seventh grade pupils on the five mental tests and also the distribution of the mental ages of one hundred twenty-five of these pupils on the Stanford-Binet, together with the median chronological age, the mean and median mental ages for each test, and the standard deviation for each distribution. It will be seen that this group is not quite "normal" if one may judge by the central tendency of the different distributions of mental ages shown in this table. If

the group were exactly normal, the median mental age on a perfectly discriminatory test would exactly coincide with the chronological age. Actually we find that the median mental age on each of the tests is above the median chronological age, ranging from seven months above,

TABLE I.—DISTRIBUTIONS OF MENTAL AGES IN MONTHS OF THE SEVENTH GRADE PUPILS IN TERMS OF FIVE INTELLIGENCE TESTS

MA	Terman	Otis	National	Haggerty	Multi-m	Binet
240–244	.....	.....	.....	.....	1	
235–239	.....	.....	.....	.....	0	
230–234	.....	.....	.....	.....	0	
225–229	.....	.....	.....	.....	0	
220–224	.....	1	.....	.....	1	
215–219	.....	1	.....	1	2	
210–214	1	4	3	3	3	
205–209	0	2	5	2	0	
200–204	1	4	3	1	1	1
195–199	4	9	8	7	4	1
190–194	1	7	8	8	6	1
185–189	5	9	3	5	5	5
180–184	8	9	1	9	12	2
175–179	12	9	3	12	8	13
170–174	13	10	12	13	13	6
165–169	13	17	8	17	11	14
160–164	25	6	23	24	18	17
155–159	22	22	29	22	13	11
150–154	22	15	21	17	16	14
145–149	32	14	25	10	19	12
140–144	18	7	14	15	17	9
135–139	10	15	10	7	13	9
130–134	3	13	9	4	10	5
125–129	.....	4	3	8	7	2
120–124	.....	6	1	0	5	0
115–119	.....	3	0	2	3	0
110–114	.....	0	1	1	1	2
105–109	.....	1	.....	0	1	1
100–104	.....	1	.....	2		
95– 99	.....	1				
Total.....	190	190	190	190	190	125
Median MA....	157.27	158.40	156.89	161.45	156.15	158.86
Median CA	149.09	149.09	149.09	149.09	149.09	149.09
S. D.....	15.35	24.8	20.35	21.15	23.15	17.3
Mean MA.....	159.4	161.1	160.3	162.78	159	158

on the Multi-mental test, to twelve months above on the Haggerty test.

To use the median chronological age as a criterion of the discriminative capacity of the intelligence tests used, it was necessary to multiply the median chronological age (one hundred forty-nine months) by the average intelligence quotient of the five tests (106.55). The tests may then be ranked for discriminating capacity according to the amount of deviation of the median mental age of each test (see Table I) from the criterion mental age (158.75 months). In terms of the extent of their deviation from the criterion, the tests would rank as follows:

1. Otis
2. Terman
3. National
4. Multi-mental
5. Haggerty

There is a marked difference in the range of mental ages for this group as determined by the different intelligence tests. The Terman test has the smallest, a range of eighty-two months, extending from one hundred thirty-two to two hundred fourteen months; the Multi-mental has the widest, a range of one hundred thirty-six months and extending from one hundred eight to two hundred forty-four months. These facts would seem to indicate that the tests vary a great deal in their ability to discriminate between the true mental ages of children.

The standard deviations of the mental age distributions show a marked difference, (Table I), especially so considering the fact of a very narrow range of difference in median and mean mental ages. Assuming that a discriminating test would have a comparatively small standard deviation of mental ages, the tests may be ranked as follows:

1. Terman
2. National
3. Haggerty
4. Multi-mental
5. Otis

There is a very small range of difference in the median mental ages of the five tests. If we consider the median mental age of the Stanford-Binet as a criterion of discrimination and validity for the five tests, they rank as follows (Table I):

1. Otis
2. National
3.5 Terman
3.5 Haggerty
5. Multi-mental

In Table II are shown the distributions of the intelligence quotients on the five group tests and also the distribution of IQ's of the one

TABLE II.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENTS OF THE SEVENTH GRADE PUPILS IN TERMS OF FIVE INTELLIGENCE TESTS

IQ	Terman	Otis	National	Haggerty	Multi-m	Binet
170–174	.....	.....	.....	.....	2	
165–169	.....	.....	.....	.....	1	
160–164	1	.....	.....	.....	0	
155–159	1	1	1	2	0	
150–154	0	4	3	0	2	
145–149	0	7	3	4	2	
140–144	0	6	5	5	2	1
135–139	7	4	9	9	8	1
130–134	8	14	8	6	8	5
125–129	13	6	7	12	14	7
120–124	7	11	8	10	9	8
115–119	19	14	7	22	12	14
110–114	18	16	19	21	20	17
105–109	24	23	26	24	15	13
100–104	30	16	21	18	16	23
95– 99	25	21	29	22	19	8
90– 94	13	13	24	8	28	10
85– 89	13	5	5	5	7	10
80– 84	8	2	7	12	9	4
75– 79	3	9	6	5	8	3
70– 74	.....	5	1	1	5	0
65– 69	.....	1	1	3	3	1
60– 64	.....	2	.....	1		
Total.....	190	190	190	190	190	190
Median IQ.....	105.62	107.39	105.19	109.58	105	106.34
Standard Deviation.....	15.35	20.85	18	18.5	20.3	14.4
Mean.....	107.57	108.7	107.9	109.55	107.05	106.7

hundred twenty-five for the Stanford-Binet together with the medians, means, and standard deviations of the distributions.

on the Multi-mental test, to twelve months above on the Haggerty test.

To use the median chronological age as a criterion of the discriminative capacity of the intelligence tests used, it was necessary to multiply the median chronological age (one hundred forty-nine months) by the average intelligence quotient of the five tests (106.55). The tests may then be ranked for discriminating capacity according to the amount of deviation of the median mental age of each test (see Table I) from the criterion mental age (158.75 months). In terms of the extent of their deviation from the criterion, the tests would rank as follows:

1. Otis
2. Terman
3. National
4. Multi-mental
5. Haggerty

There is a marked difference in the range of mental ages for this group as determined by the different intelligence tests. The Terman test has the smallest, a range of eighty-two months, extending from one hundred thirty-two to two hundred fourteen months; the Multi-mental has the widest, a range of one hundred thirty-six months and extending from one hundred eight to two hundred forty-four months. These facts would seem to indicate that the tests vary a great deal in their ability to discriminate between the true mental ages of children.

The standard deviations of the mental age distributions show a marked difference, (Table I), especially so considering the fact of a very narrow range of difference in median and mean mental ages. Assuming that a discriminating test would have a comparatively small standard deviation of mental ages, the tests may be ranked as follows:

1. Terman
2. National
3. Haggerty
4. Multi-mental
5. Otis

There is a very small range of difference in the median mental ages of the five tests. If we consider the median mental age of the Stanford-Binet as a criterion of discrimination and validity for the five tests, they rank as follows (Table I):

1. Otis
2. National
3.5 Terman
3.5 Haggerty
5. Multi-mental

In Table II are shown the distributions of the intelligence quotients on the five group tests and also the distribution of IQ's of the one

TABLE II.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENTS OF THE SEVENTH GRADE PUPILS IN TERMS OF FIVE INTELLIGENCE TESTS

IQ	Terman	Otis	National	Haggerty	Multi-m	Binet
170–174	.....	.....	.....	.....	2	
165–169	.....	.....	.....	.....	1	
160–164	1	.....	.....	.....	0	
155–159	1	1	1	2	0	
150–154	0	4	3	0	2	
145–149	0	7	3	4	2	
140–144	0	6	5	5	2	1
135–139	7	4	9	9	8	1
130–134	8	14	8	6	8	5
125–129	13	6	7	12	14	7
120–124	7	11	8	10	9	8
115–119	19	14	7	22	12	14
110–114	18	16	19	21	20	17
105–109	24	23	26	24	15	13
100–104	30	16	21	18	16	23
95– 99	25	21	29	22	19	8
90– 94	13	13	24	8	28	10
85– 89	13	5	5	5	7	10
80– 84	8	2	7	12	9	4
75– 79	3	9	6	5	8	3
70– 74	.....	5	1	1	5	0
65– 69	.....	1	1	3	3	1
60– 64	.....	2	.....	1		
Total.....	190	190	190	190	190	190
Median IQ.....	105.62	107.39	105.19	109.58	105	106.34
Standard Deviation.....	15.35	20.85	18	18.5	20.3	14.4
Mean.....	107.57	108.7	107.9	109.55	107.05	106.7

hundred twenty-five for the Stanford-Binet together with the medians, means, and standard deviations of the distributions.

Assuming that the median intelligence quotient of the Stanford-Binet may serve as a criterion in determining the discrimination and validity of the five tests, they may be ranked according to the deviation of the median IQ's from it as follows:

1. Terman
2. Otis
3. National
4. Multi-mental
5. Haggerty

If we use the size of the standard deviation as a criterion again, and assume that the smaller the standard deviation the more discriminating, as some have assumed,[1] the tests rank as follows (Table II):

1. Terman
2. National
3. Haggerty
4. Multi-mental
5. Otis

Table III presents the inter-correlations between the mental ages obtained by the different tests.

TABLE III.—INTER-CORRELATIONS BETWEEN MENTAL AGES OF ONE HUNDRED NINETY SEVENTH GRADE PUPILS AS DETERMINED BY FIVE INTELLIGENCE TESTS: COEFFICIENTS ALL POSITIVE

Name of test	Terman	Otis	National	Haggerty	Multi-mental
Terman	....	.785	.742	.779	.766
Otis	....	....	.821	.751	.687
National	....	....	....	.785	.675
Haggerty	....	....	....	....	.712
Average	.753	.761	.755	.756	.695

PE's range from .015 to .025.

If the extent of the inter-correlation of mental ages indicate their degree of validity, they rank as follows:

[1] Sangren, Paul: *Op. cit.*

1. Otis
2. Haggerty
3. National
4. Terman
5. Multi-mental

Using the intelligence quotients in the same way as the mental ages were used in Table III the tests rank as follows:

1. Haggerty
2. Otis
3. Terman
4. National
5. Multi-mental

The foregoing tables and other data will serve to make clear the final table which presents the ranking of each of the five tests according to each of twenty-nine criteria used, together with the total (or average) of the rankings so obtained. A further word of explanation may be necessary regarding some of the criteria. Thus criteria twelve and thirteen were set up by first transmuting the mental ages and IQ's of the five tests into equivalents of the Stanford Binet by using the standard deviation technique suggested by Miller.[1] The average deviation of these equivalents for each test from the composite of all was then determined.

Probably not all the criteria are of equal value. No attempt has been made to study the data from the standpoint of criteria assumed to be of more value than others, except that criteria 2 and 5, based on the size of the standard deviation, are perhaps the most unsatisfactory of all. If they are eliminated, some difference in the total ranking of each test occurs, and the rank order of the test is slightly changed also.

From Table IV it will be seen that the tests rank on the average of 29 criteria as follows:

1. Otis
2. Terman
3. Haggerty
4. National
5. Multi-mental

[1] Miller, W. S.: Variation and Significance of IQ's Derived from Group Test. *Journal of Educational Psychology.* 1925.

TABLE IV.—TABULAR RANKING OF TESTS FOR DISCRIMINATION AND VALIDITY WHEN CHECKED AGAINST TWENTY-NINE CRITERIA

Criteria	Terman	Otis	National	Haggerty	Multi-m
1. Deviation of median MA	2	1	3	5	4
2. Standard deviation of MA's	1	5	2	3	4
3. Deviation of median MA's from S-B median MA	3.5	1	2	3.5	5
4. Deviation of median IQ's from S-B median IQ	1	2	3	5	4
5. Standard deviation of IQ's	1	5	2	3	4
6. Inter-correlation of MA's	4	1	3	2	5
7. Inter-correlation of IQ's	3	2	4	1	5
8. Average deviation of raw MA's from S-B MA	1	5	2	3	4
9. Average deviation of raw MA's from composite MA	1	5	3	2	4
10. Average deviation of raw IQ's from S-B IQ	1	5	2	3	4
11. Average deviation of raw IQ's from composite IQ	1	5	2	3	4
12. Average deviation of MA's (S-B equivalents) from composite MA	5	4	1	2	3
13. Average deviation of IQ's (S-B equivalents) from composite IQ	3	5	2	1	4
14. Correlation of MA's and letter grades in reading	3	1	2	5	4
15. Correlation of MA's and letter grades in arithmetic	3	1	4	2	5
16. Correlation of MA's and letter grades in social science	3	2	5	1	4
17. Correlation of IQ's and letter grades in reading	1	2	3	4	5
18. Correlation of IQ's and letter grades in arithmetic	3	1	4	2	5
19. Correlation of IQ's and letter grades in social science	3	1.5	4	1.5	5
20. Correlation of MA's and paragraph meaning scores	4.5	1	4.5	2	3
21. Correlation of MA's and arithmetic reasoning scores	1	3	4	2	5
22. Correlation of IQ's and paragraph meaning scores	3	2	4	1	5
23. Correlation of IQ's and arithmetic reasoning scores	3	1	4	2	5
24. Correlation of MA's and composite judgment of teachers	2	1	5	3	4
25. Correlation of IQ's and composite judgment of teachers	3	1	4	2	5
26. Multiple correlation of MA's on pairs of tests and composite judgment of teachers	2	1	5	3.5	3.5
27. Multiple correlation of IQ's on pairs of tests and composite judgment of teachers	3	1	4	2	5
28. Correlation of MA's with average semester letter grades in three subjects	3	1	4	2	5
29. Correlation of IQ's with average semester letter grades in three subjects	2.5	1	4	2.5	5
Totals	70.5	67.5	95.5	74	127.5
If the two criteria (Nos. 2 and 5) involving standard deviation are omitted, the totals are:	68.5	57.5	91.5	68	119.5

It will be seen that three of the tests are very nearly equal in value as judged by the total (or average) ranking; namely, Otis, Terman, and Haggerty. In so far as their value for this particular school level is concerned they are practically interchangeable. The National is considerably less adequate and the Multi-mental much less adequate.

SUMMARY

Five group tests were administered to one hundred ninety 7B pupils. Twenty-nine criteria of discrimination and validity were applied to the data obtained from these tests. For each criterion the tests were ranked in order, and for the average or total of the twenty-nine a final ranking was obtained. This final ranking places the tests in the following order:

1. Otis
2. Terman
3. Haggerty Delta 2
4. National
5. Multi-mental

The first three are shown to be of almost equal value as judged by the total ranking of the twenty-nine criteria. The National is of considerably less, and the Multi-mental of much less value by the same criterion.

ADDITIONAL DATA ON THE RELIABILITY AND VALIDITY OF PHOTOGRAPHIC EYE-MOVEMENT RECORDS[1]

ALVIN C. EURICH

University of Minnesota

In the February issue of the Journal, the writer published data concerning the reliability and validity of photographic eye-movement records of college students. In view of the relatively high reliability and low validity according to the criteria of reading, achievement, and intelligence tests, arrangements were made to check the results with a group of elementary school pupils.

During the spring of 1932, slightly more than one hundred pupils in grades 4A, 5B, and 5A of the Tuttle School in Minneapolis were transported to the University of Minnesota Campus in groups of four to seven, for the purpose of securing the photographic eye-movement records. Each pupil read the following three paragraphs.

I. "He had in his house a housekeeper past forty, a niece under twenty, and a lad for the field and market place, who used to saddle the hack as well as handle the pruning hook. The age of this gentleman of ours was bordering on fifty; he was of a hardy habit, spare, gaunt featured, a very early riser and a great sportsman."

II. "Ants live in hills or nests made of sand or dirt. Many ants work together to make one nest or ant city. One ant hill has often as many as two hundred houses. A nest is sometimes a tunnel in the earth and sometimes a large hill with hallways under the ground."

III. "They went across the hall to a door at the back of the house. It opened before them and disclosed a long, bare, melancholy room, made barer still by lines of desks. At one of these a lonely boy was reading near a feeble fire; and Scrooge sat down upon a form, and wept to see his poor forgotten self as he had used to be."

The first and third paragraphs were printed in 12 point type while the second was in 14 point type. Approximately one-half of the group read the paragraphs in 1-2-3 order while the other half read them in 3-2-1 order. In addition the pupils were asked four questions on each paragraph immediately after reading it in order to ascertain whether they understood what they had read. This comprehension score served as a criterion for estimating the validity of the eye-movement

[1] This study was subsidized by the Graduate School of the University of Minnesota. The writer also gratefully acknowledges his indebtedness to Alice Albert Eurich who photographed the eye-movements of the entire group and to Mr. H. C. Cooper, principal of the Tuttle School, for his willingness to coöperate.

records. Other criteria consisted of the following measures that had been given to the children shortly before or immediately after the eye movement records had been obtained: Chapman-Cook Speed of Reading Test, Forms A and B; Pressey Diagnostic Reading Tests and the Pressey Intermediate Classification Test.

The coefficients of reliability for the number of fixations, number of regressions and average duration of fixations as derived from the photographic records are assembled in Table I. Only those children whose records were complete on all three paragraphs were included in these calculations. The coefficients represent the relationship between the measures on paragraphs 1 and 2, 1 and 3, 2 and 3, and the estimated

TABLE I.—RELIABILITY OF EYE-MOVEMENT RECORDS (N = 80)

Measure	Paragraphs	r PE
Number of fixations	1 and 2	.70 ± .04
	1 and 3	.74 ± .03
	2 and 3	.87 ± .02
Estimated	All paragraphs	.91 ± .01
Number of regressions	1 and 2	.68 ± .04
	1 and 3	.77 ± .03
	2 and 3	.87 ± .02
Estimated	All paragraphs	.91 ± .01
Average duration of fixations	1 and 2	.69 ± .04
	1 and 3	.67 ± .04
	2 and 3	.62 ± .05
Estimated	All paragraphs	.86 ± .02

reliability of the measures derived from a combination of the three paragraphs, the estimation being made by the Douglass-Cozens formula[1] for the reliability of test batteries. It is clearly evident that for fixations and regressions, the reliability of paragraphs 1 and 2 is the lowest; of paragraphs 1 and 3, somewhat higher; and of paragraphs 2 and 3, relatively high. When the three paragraphs are combined, the reliability coefficients are as high as for the best tests available at the present time. For the average duration of fixations the reliability as determined from any two paragraphs does not reach as high a point as for fixations and regressions. The estimated reliability on the three

[1] Douglass, H. R. and F. W. Cozens: Reliability of Test Batteries. *Journal of Educational Psychology*, Vol. XX, 1929, pp. 369–377.

paragraphs combined, however, is entirely satisfactory. These results are wholly in accord with those derived from the group of college students.

The inter-correlations of the eye-movement records are presented in Table II. The product-movement coefficients between the number of fixations and number of regressions range from .81 to .94; all are higher than the relationship that Tinker[1] has reported for a comparable group (.74 ± .04) and lead to the deduction that the number of fixa-

TABLE II.—INTER-CORRELATION OF EYE-MOVEMENT RECORDS

	Number of regressions	Average duration of fixations
Paragraph I		
Number of fixations	.84 ± .02	.38 ± .06
Number of regressions	.....	.22 ± .06
Paragraph II		
Number of fixations	.87 ± .02	.19 ± .07
Number of regressions	.....	.10 ± .07
Paragraph III		
Number of fixations	.94 ± .01	.38 ± .06
Number of regressions	.....	.41 ± .06
Total paragraphs I, II, III		
Number of fixations	.81 ± .02	.36 ± .06
Number of regressions	.....	.36 ± .06

tions and the number of regressions are measures of the same ability. Since the number of regressions is so much smaller than the total number of fixations, it would be a decided economy to use only the former measure.

The coefficients of correlation ranging from .19 to .38 for the number of fixations and the average duration of fixations are closely in agreement with the coefficient of .33 reported by Tinker. Similarly the correlation of .10 to .41 between the number of regressions and the average duration of fixations do not deviate more than four probable

[1] Tinker, M. A.: "Photographic Measures of Reading Ability." *Journal of Educational Psychology*, Vol. XX, 1929, pp. 184–191.

errors from the relationship previously reported (.19).[1] The evidence is uniform and clear in pointing to the fact that the average duration of fixations is a different measure from either the number of fixations or the number of regressions.

The data in Table III represent the relationship between the photographic eye-movement records and various reading and intelligence

TABLE III.—THE RELATION OF PHOTOGRAPHIC EYE-MOVEMENT RECORDS TO READING AND INTELLIGENCE TESTS

	Paragraphs	Chapman-Cook A and B	Pressey diagnostic reading		Comprehension scores	IQ Pressey classification intermediate
			Vocabulary	Paragraph reading		
Number of fixations.	1	−.53 ± .05	−.18 ± .08	−.28 + .08	−.15 ± .07	−.35 ± .06
	2	−.40 ± .06	.....	.....	.....	−.37 ± .06
	3	−.36 ± .07	.....	.....	.....	−.37 ± .06
	1, 2, 3	−.43 ± .06	−.28 ± .08	−.27 ± .08	−.17 ± .07	−.38 ± .06
Number of regressions.	1	−.42 ± .06				
	2	−.55 ± .05				
	3	−.34 ± .07				
	1, 2, 3	−.41 ± .06	−.26 ± .08	−.23 ± .08	−.11 ± .07	−.32 ± .06
Average duration of fixation.....	1	−.28 ± .07	.....	.....	.....	−.35 ± .06
	2	−.31 ± .07	.....	.....	.....	−.35 ± .06
	3	−.25 ± .07	.....	.....	.....	−.34 ± .06
	1, 2, 3	−.34 ± .06	−.15 ± .08	−.15 ± .08	−.16 ± .07	−.42 ± .06
Pressey IQ.....	.....	.56 ± .06				

tests. With the Chapman-Cook Speed of Reading Tests, Forms A and B combined, the coefficients range from −.25 to −.55. Since the reliability of the Chapman-Cook test is approximately .91 and the reliability of the photographic measures about the same when the three paragraphs are combined, it is evident that the low relationship between the two is not due wholly to the unreliability of the measures. The remaining relationships in the table are even lower. Again it may be emphasized that when photographic eye-movement records are evaluated in terms of reading tests for which the reliability and validity have been demonstrated the photographic records do not appear to be highly satisfactory. Whether the records obtained from the reading of long selections or after the subject has been given several practice periods would be more satisfactory remains to be determined.

[1] Tinker: *Ibid.*

SUMMARY

1. Photographic eye-movement records of reading for elementary school pupils are approximately as reliable as the best achievement and intelligence tests available. These results coincide with those derived from college students.

2. The inter-relationships of the eye-movement records show that the number of regressions and the number of fixations are measures of the same ability whereas the average duration of fixations is not highly related to either of the other measures.

3. While the validity of photographic eye-movement records for elementary school pupils as determined by various reading and intelligence tests is somewhat higher than for college students, it is not sufficiently so as a measure of rate or comprehension in reading to be regarded as highly satisfactory for diagnostic purposes.

AN EVALUATION OF CERTAIN ASPECTS OF A PROGRAM OF CHARACTER EDUCATION

D. D. FEDER AND L. W. MILLER

University of Denver

1. PROBLEM

This study has attempted to evaluate certain phases of a comprehensive program of character education for boys. This program is carried on by an organization in Denver, Colorado, which, for the purposes of this paper, will be known as the "X" organization.

Dependent upon their age at entrance, boys are given anywhere from two to four or five years of training, which consists chiefly of military training, and an intensive program of optional club work. The boys are allowed to make their club selections according to hobbies, crafts, cultural activities, etc.

It is the claim of the organization that the training it gives instils in boys "functioning Christian character." The creed, military training, and military trappings do not, claim the officials of this organization, contribute to any more favorable feeling toward war.

The following quotation from the literature of the organization best serves to outline its aims and objectives:

> To a ("X") boy, his parents, close friends, or careful observers of the lives of these boys, the ("X") uniform has come to represent, in addition to these (correct posture, neatness, obedience, respect, and efficiency), other certain very definite characteristics that are most desirable in the growing lad. These characteristics are best explained in the precepts of the organization which become part of every "X" boy's life: Be Kind, Live Pure, Speak Truth, Right Wrong, Defend the Weak, and Play the Game Square. Such abstract ideas become effective only when they are repeatedly experienced and become habits of life.

In the light of these claims this study has attempted to ascertain whether, when there are composite pictures of two groups of boys, with emphasis placed upon various traits of character and behavior, it is possible to tell which group has had the ("X") training and which has not. In order to do this, an attempt was made to evaluate the work as regards:

1. Knowledge of health facts and habits;
2. Citizenship, as expressed in knowledge of the elements of good citizenship, and as manifested in behavior;
3. Attitudes toward war and peace—(it has been charged that this organization encourages a feeling favoring war, due to its militaristic régime and trappings);

4. Emotional balance, as manifested in feelings of superiority and inferiority;
5. General factors of school and social behavior.

2. METHOD AND PROCEDURE

With the coöperation of the officials of the Denver Public Schools, boys in East and South high schools were used in this study. Two groups were employed. There were one hundred boys in each group. One was made up of boys who had received a full course of the ("X") training, but no training as Boy Scouts, Cadets, etc. The other group contained boys who had received no formal organizational training of any kind. The boys in the second group were matched on the basis of intelligence with those in the first. Factors of age, school grade, junior high school attended, and socio-economic status were held relatively constant. Thus the strongest variable present was the factor of training, the effect of which this study sought to determine.

In line with the five types of information sought, the following instruments were used:

1. Gates-Strang Health Knowledge Test
2. Good Citizenship Test (Hartshorne and May)
3. War-Peace Attitude Scale (Thurstone and Droba)
4. Emotional Attitude Test (Feder)
5. Haggerty-Olson-Wickman Behavior Rating Schedule

Other information secured related to the amount of absence and tardiness, taken for the semester during which the study was carried on. The first four instruments, plus the Otis S-A Test of Mental Ability (Advanced Form *A*), were administered directly to the subjects. The Behavior Rating Schedules were filled out by teachers of the boys.

3. RESULTS AND CONCLUSIONS

In Table I p. 387, the means, standard deviations, SD of differences between the means, and the ratios of $\frac{D}{SD_{diff.}}$ of the two groups on the various instruments used and information secured, are given.

Inspection of these results and evaluation of mean differences, using the formula devised by Lindquist and Wilks of the University of Iowa,[8,13], reveals that there are no statistically significant differences.

On the basis of the results secured by this study, the following conclusions regarding the work of the ("X") organization seem warranted:

TABLE I.—MEANS, STANDARD DEVIATIONS, $SD_{diffs.}$, $\frac{D}{SD_{diff.}}$, ON TESTS USED, BEHAVIOR RATING SCHEDULES, AND OTHER INFORMATION, TRAINED AND UNTRAINED GROUPS

Test	Group	Mean	SD	$SD_{diff.}$	$\frac{D}{SD_{diff.}}$
Otis S.A. (Advanced)	Trained	54.09	8.904	1.103	.108
Otis S.A. (Advanced)	Untrained	53.97	6.523	1.103	.108
IQ	Trained	114.42	9.390	1.293	.116
IQ	Untrained	114.27	8.880		
Good Citizenship	Trained	75.99	7.350	.875	1.954
Good Citizenship	Untrained	77.70	7.260		
Emotional Attitude	Trained	8.58	9.780	3.864	.551
Emotional Attitude	Untrained	10.71	8.850		
War-Peace Attitude	Trained	4.21	.729	.105	.285
War-Peace Attitude	Untrained	4.18	.858		
Health Knowledge	Trained	54.32	3.840	.391	.562
Health Knowledge	Untrained	54.10	3.440		
Behavior Rating (A)	Trained	15.02	15.896	2.096	.381
Behavior Rating (A)	Untrained	14.22	14.760		
Behavior Rating (B)	Trained	72.00	17.845	2.599	.519
Behavior Rating (B)	Untrained	73.35	19.950		
Absences	Trained	4.63	4.92	.843	2.168
Absences	Untrained	6.46	6.84		
Tardinesses	Trained	3.25	4.28	.562	.160
Tardinesses	Untrained	3.34	3.64		
Scholastic Average	Trained	75.65	8.050	.857	.466
Scholastic Average	Untrained	75.25	7.750		
Age (in months)	Trained	198.16	13.20	1.96	1.102
Age (in months)	Untrained	200.32	14.40		

1. The training does not give boys better knowledge concerning the fundamentals of good citizenship;

2. There is no evidence that the trained boys behave better in school and social situations as a result of their training;

3. The training does not eliminate personality-behavior problems;

4. The training does not give boys any especial attitude of superiority or inferiority;

5. The training does not give boys an added stimulus to do good work in their scholastic subjects;

6. The training does not give boys better knowledge concerning health habits and practices;

7. The training does not give boys a more favorable attitude toward war. The trained group, as measured by the Thurstone-Droba War Attitude Scale is not more militaristic than the untrained, measured by the same scale;

8. Where observation of certain rules becomes habitual and implicit, such as in the case of absence from school, the training seems to function in making boys more regular in school attendance. In the matter of tardiness the training seems to make no difference.

4. SUPPLEMENTARY CONCLUSIONS

Certain other results of a more general educational significance have also come to light as a result of this study. These are given in the following tables:

TABLE II.—CORRELATIONS WITH INTELLIGENCE

Test or scale	r	PE$_r$
Emotional Attitude	.142	.047
Health Knowledge	.425	.039
Good Citizenship	.282	.044
War-peace Attitude	.053	.047
Scholastic Average	.412	.040
Behavior Rating (A)	.257	.045
Behavior Rating (B)	.238	.045

TABLE III.—CORRELATIONS WITH AGE

Test or scale	r	PE$_r$
Emotional Attitude	−.020	.048
Health Knowledge	−.005	.047
Good Citizenship	.071	.047
War-peace Attitude	.037	.046
Behavior Rating (A)	−.105	.047
Behavior Rating (B)	−.100	.047

According to prevalent modes of interpretation, the only two significant coefficients indicated in Table II are those between intelligence and the Health Knowledge scores, (.425), and intelligence and scholastic

TABLE IV.—CORRELATIONS WITH SCHOLASTIC AVERAGE

Test or scale	r	PE_r
Good Citizenship	.141	.047
Behavior Rating (A)	.566	.032
Behavior Rating (B)	.458	.039
Intelligence	.412	.040

TABLE V.—CORRELATIONS WITH GOOD CITIZENSHIP

Test or scale	r	PE_r
Intelligence	.282	.044
Age	.071	.047
Scholastic Average	.141	.047
Behavior Rating (A)	.147	.047
Behavior Rating (B)	.070	.047

average, (.412). It is quite possible, of course, that the Health Knowledge test contains such items as make it a fairly good intelligence test. This conclusion seems warranted in view of the higher correlation here than with any other of the testing instruments used. That scholastic average is related to intelligence has long been an established fact. As a matter of fact, somewhat higher correlations than this have been reported, some of them ranging to .60.

Scores on the Behavior Rating Schedules were correlated with intelligence quotients, to test the assumption that increase in intelligence is accompanied by better understanding of ethical relations, with the result therefore that behavior should become increasingly better in children with higher intelligence. That this assumption is questionable is indicated by the very low correlations, (.257 and .238). However, since there is some correlation present, it may be indicative of the presence of relationship, and that if the instruments of measurement, as well as the raters, were more accurate, the correlation might be higher and more significant.

With the exception of the Behavior Rating Schedules, there are practically zero correlations of the various factors with age, as shown by Table III. In the case of the behavior problem rating, it seemed reasonable to expect that with increasing age and understanding, behavior problems should decrease. This would seem consistent in the light of some modern theories of character which hold that with increase in age there comes about an integration of habit patterns, so that,

apparently, the personality is guided by a single principle or set of principles. Again the assumption was found to be invalid, the results indicating no such coördination or integration. As a matter of fact, the negative correlations of −.105 and −.100 would hint that as the boys grow older their behavior problems tend to increase. This would suggest that as responsibilities and problems become more complex with increasing age, behavior difficulties are found more frequently.

Table IV gives the correlations between scholastic average and other ratings. Those of chief interest are the behavior problem scores, as correlated with scholastic average. The correlation of .566 between Part A (the global rating on general problems) and intelligence suggests several interesting possibilities. In the first place, the unreliability of teachers' grades has been demonstrated over and over. Especially has this been shown where the teacher does not know whose work he is grading. Therefore, it might be assumed that teachers' grades are influenced markedly by the "halo" effect which Newcomb[9] reports in his experiment in behavior prediction. However, there is another aspect to the problem, namely, that the pupil who works hardest in school has the least amount of time to give trouble. Dependent upon one's prejudices one may choose either of these explanations or perhaps suggest others. However, it seems to the writers that there is a preponderance of data favoring the first explanation, and not so much in favor of the second.

Table V gives the correlations between knowledge of good citizenship and behavior. Although the latter scale does not measure specifically good citizenship, there are a large number of related items present. The results would indicate somewhat the same thing that Hartshorne and May[7] found, namely, that there is little or no correlation between moral knowledge and moral action.

The conclusions that seem justified from the data presented in Tables II to V inclusive, are as follows:

1. There is no significant correlation between emotional attitude scores and intelligence, or age;

2. The correlation between health knowledge and intelligence would indicate that this test measures to some degree intelligence, and this is further verified by the lack of correlation between this test and age;

3. There is a low correlation between good citizenship knowledge and intelligence, and there is a still lower, in fact, negligible correlation with age;

4. There is practically zero correlation between scores on the war-peace attitude scale and intelligence and age;

5. The presence of the intelligence factor in determining school grades is indicated by the correlation found. Also there is a possibility that school grades are influenced by the behavior of the pupils in their classrooms;

6. There is a slight indication that with increasing intelligence behavior problems decrease. With increasing age there is a very slight tendency for behavior problems to increase. These correlations are so low, however, that for the present, much confidence cannot be placed in them;

7. As measured by the instruments here employed there seems to be no relationship between knowledge of the elements of good citizenship and behavior.

BIBLIOGRAPHY

1. Feder, Daniel D.: "A Preliminary Report of an Attempt to Devise a Test for the Diagnosis of Inferiority and Superiority Complexes." (As yet unpublished.)
2. ———: "The Emotional Attitude Test." (As yet unpublished.)
3. Gates, Arthur I., and Ruth Strang: "Health Knowledge Test and Manual." Bureau of Publications, Teachers College, Columbia University, 1925.
4. Haggerty, Melvin E., Willard C. Olson, and E. Koster Wickman: "Behavior Rating Schedules and Manual." World Book Company, 1930.
5. Hartshorne, Hugh, and Mark A. May: "Studies in Deceit." Chapters 18, 19, 20. The Macmillan Company, 1928.
6. ———: "Studies in Service and Self-Control." Chapters 12, 27. The Macmillan Company, 1929.
7. Hartshorne, Hugh, Mark A. May, and Frank K. Shuttleworth: "Studies in the Organization of Character." The Macmillan Company, 1930, pp. 498.
8. Lindquist, Everett F.: "The Significance of a Difference Between 'Matched' Groups." *Journal of Educational Psychology*, Vol. XXII, March, 1931, pp. 197–204.
9. Newcomb, Theodore: "An Experiment Designed to Test the Validity of a Rating Technique." *Journal of Educational Psychology*, Vol. XXII, April, 1931, pp. 279–289.
10. Otis, Arthur S.: "Self-Administering Tests of Mental Ability and Manual." World Book Company, 1928.
11. Thurstone, Louis L. and E. J. Chave: "The Measurement of Attitude." University of Chicago Press, 1929, pp. 96.
12. Thurstone, Louis L. and D. D. Droba: "War-Peace Attitude Scales." University of Chicago Press, 1931.
13. Wilks, Samuel S.: "The Standard Error of the Means of Matched Samples." *Journal of Educational Psychology*, Vol. XXII, March, 1931, pp. 205–208.

A SURVEY OF AN OPPORTUNITY ROOM FOR GIFTED CHILDREN

NORMA V. SCHEIDEMANN

University College, University of Southern California

AND

MARGARET S. SMITH

Los Angeles City Schools

One of the most recent movements of our public schools is the segregation of gifted children for specialized instruction. In most schools the selection of this group of children is based solely on a high intelligence quotient. Children doing exceptional work, but who have an average intelligence quotient ordinarily are not admitted since it is thought that their attainments are accomplished by excessive effort. But children of high intelligence ratings doing average or even inferior work are admitted readily since it is thought that the regular methods of instruction are not adapted to gifted children.

Many studies of gifted children lead us to expect gifted children to be superior in most desirable qualities so that we are prone to think of the members of an opportunity room for gifted children to be superior physically, emotionally, scholastically, as well as intellectually. Interest in the question as to whether this condition obtained in small groups of gifted children led the present writers to make a survey of an unselected opportunity room for gifted children in a representative city school system.[1] This room is an ungraded room of twenty-two children doing from 4A to 6A work.

The children of this room were tested in various fields by means of tests readily available to the average class-room teacher. The intelligence tests were administered by the psychology department of the school system. All other tests were given by the teacher in charge of the children. The accompanying table shows the aggregated ratings attained by these children. The intelligence ratings are based on the various tests specified. Hearing was tested by the whisper test, an admittedly inadequate means of measurement, but the only test available. Visual acuity was tested by the Snellen test charts; and posture by the customary standards accepted by the United States

[1] Opportunity A, Miles Avenue School, Huntington Park, Los Angeles City Schools. Second semester of school year 1931–1932.

Department of Labor, Children's Department. Eyedness was tested by the paper-hole test. The ratings in scholarship are those attained in the Stanford Achievement Tests. The other ratings are the personal judgments of the teacher of the group, who has had extended experience in teaching both gifted and normal children, and who has had training in diagnosing various types of exceptionality.

This table shows that defective hearing and defective vision is quite common among this particular group. Not a single child's posture could be rated "excellent," and only slightly more than half have good posture. Five children found to be left-eyed are right-handed writers. Their teacher stated that while as a whole these children are very pleasant and a delightful group with which to work, still a critical and unbiased evaluation of their emotional stability resulted in describing only two of these children as emotionally normal and stable. The heights and weights of these children are not given because of the difficulty of comparing them with the norms of various builds for the many ages found among the group. In regard to athletic activities the teacher in charge says: "Athletic achievement is painfully lacking with my group of children. I have had one girl who won the district championship in running, and two or three boys who have placed in track events, but that is all. They are interested in school games and usually there are one or two on the different teams; not nearly the amount I would like, however." In the motor control of finer muscles, as in writing, this group is decidedly below par.

No illustration of the type of work that these children do in literary fields is given here. It must be emphasized, however, that much of their creative work in poetry and short story writing is of an unbelievably high grade of excellence. The conclusion that the teacher of this group of gifted children has derived is that intellectually and scholastically this group is far superior to normal children, but in physical and emotional qualities these children compare unfavorably with normal children.

The results of this study show us that children gifted intellectually may be average or even inferior in other traits. General statements based on the statistical findings of large groups of gifted children cannot be applied to the individual gifted child nor to the small group of gifted children. Each gifted child must be studied as an individual; his weak traits must be discovered and trained as well his strong traits. Many times the training of the gifted child's weak traits is ultimately of greater value than the disproportionate training of his gifted traits.

TABLE I.—INTELLECTUAL, SCHOLASTIC, PHYSICAL AND EMOTIONAL RATINGS OF GIFTED CHILDREN IN AN OPPORTUNITY ROOM

	Grade	Chronological age	Terman IQ	Haggerty IQ	National IQ	Reading: Paragraph—years acceleration for age	Reading: Word—years acceleration for age	Spelling—years acceleration for age	Arithmetic acceleration for age: Reasoning	Arithmetic acceleration for age: Computation	Writing	Total educational acceleration in years[1]	Hearing: Right ear	Hearing: Left ear	Vision: Right eye	Vision: Left eye	Posture	Articulation	Eyedness	Hand used in writing	Emotional adjustment
B. T.	6A	11.3	152	164	...	6.3	5.5	4.8	2.2	5.0	Good	5.9	29/20	29/20	19/10	19/10	Good	Good	R	R	Sensitive. Nervous pushing of lower lip by tongue. Not popular with group. No sense of rhythm. Craves love. Has alibi for everything. Father is doctor.
K. M.	6A	9.3	153	174	145	6.8	5.9	4.8	5.3	6.9	Poor	5.4	29/20	15/20	5/10	5/10	Good	Fair	R	R	Unsocial; supersensitive; solitary, self-sufficient. Cares nothing for companionship of others. Father was Stanford honor student. Slight hesitancy in speech.
B. M.	6A	9.2	151	188	144	6.0	3.11	5.4	6.2	6.10	Fair	5.0	15/20	19/20	9/10	5/10	Poor	Fair	R	R	Supersensitive; unsocial. Very small. Looks underweight and malnourished. Drags feet. Father is native Italian; large family.
B. D.	4B	8.10	151	148	...	4.3	3.2	2.11	2.3	1.11	Poor	2.11	29/20	29/20	19/10	19/10	Fair	Good	R	R	Flighty; avoids play; unstable.
J. M.	4A	8.5	150	...	...	3.0	3.6	2.5	0.7	2.1	Poor	2.5	15/20	29/20	19/10	19/10	Fair	Good	R	R	Unsocial; hysterical; malacious. Both parents are teachers and have pushed child. Child a problem in every school. Bad temper and crying spells if crossed. Will sit for hours doing nothing. Looks like six year old child.
J. F.	6A	11.7	147	141	150	3.9	5.7	5.4	2.6	4.2	Fair	3.10	5/20	19/20	19/10	19/10	Good	Good	L	R	Solitary; seclusive; tormentative; rough. Over weight. Poor heart. Lazy. Gets into trouble with all yard duty teachers. Father is doctor.
L. J.	4A	9.3	145	176	...	5.1	3.8	2.5	2.2	1.1	Fair	3.3	15/20	15/20	9/10	5/10	Good	Good	R	R	Sensitive.
B. K.	4A	9.3	145	141	...	1.5	3.3	2.2	2.1	0.11	Poor	1.4	15/20	15/20	19/10	19/10	Poor	Good	R	R	Unstable. Very poor home life. Blank look most of the time. Never hears directions or knows what to do after they are explained. In a fog most of time. Father dead. Mother trying to raise family.
H. H.	5A	10.8	143	...	181	5.3	4.8	5.10	2.9	4.4	Fair	3.4	29/20	15/20	19/10	19/10	Fair	Excellent	R	R	Unsocial. Does not participate in class work. If asked simply says "No." Throws self in loose-jointed manner. Always falling down. Often described as queer. From large family—only gifted one.

TABLE I.—*Continued*

TABLE I.—*Continued*

	Grade	Chronological age	Terman IQ	Haggerty IQ	National IQ	Reading: Paragraph—years acceleration for age	Reading: Word—years acceleration for age	Spelling—years acceleration for age	Arithmetic acceleration for age: Reasoning	Arithmetic acceleration for age: Computation	Writing	Total educational acceleration in years[1]	Hearing: Right ear	Hearing: Left ear	Vision: Right eye	Vision: Left eye	Posture	Articulation	Eyedness	Hand used in writing	Emotional adjustment
B. T.	6B	10.6	142	...	150	4.2	5.2	3.5	1.8	3.1	Excellent	4.8	20/20	10/20	9/10	5/10	Good	Excellent	L	R	Very normal.
B. Te.	4A	9.1	140	...	...	2.10	3.6	5.0	2.6	1.5	Excellent	2.11	20/20	20/20	10/10	10/10	Good	Good	R	R	Sensitive.
M.. B.	5A	11.0	137	130	130	3.4	2.3	2.7	1.7	1.4	Excellent	2.11	20/20	15/20	10/10	10/10	Good	Fair	R	R	High temper. Youngest in family; others grown.
R S.	6A	9.10	135	172	152	5.10	5.6	4.3	4.10	7.8	Poor	4.9	20/20	20/20	10/10	10/10	Poor	Good	R	R	Egotistical, aggressive; excitable; unstable. Neurotic; never faces facts. Does not know how to play. Prefers to be with adults. Very untidy. Sunny disposition. Unpopular. Adopted by an solicitous woman. Very small. Looks like 6 or 7.
B. C.	6A	11.4	135	164	...	4.10	4.7	1.3	3.4	5.10	Fair	3.2	5/20	5/20	10/10	7/10	Fair	Good	R	R	Very social.
E. B.	6A	10.5	133	130	130	4.9	4.1	4.9	4.6	6.7	Excellent	3.5	15/20	20/20	10/10	10/10	Fair	Good	R	R	Tormentative.
R. T.	5A	10.9	132	137	148	2.4	2.2	1.0	3.0	3.7	Poor	1.11	15/20	15/20	10/10	10/10	Fair	Poor	L	R	Tormentative.
B. B.	6A	11.11	131	150	149	4.7	4.7	2.2	1.6	3.8	Fair	3.1	15/20	20/20	10/10	10/10	Good	Good	...	R	Unsocial.
J. N.	6B	11.2	129	148	138	1.6	5.0	1.6	2.3	4.2	Poor	3.4	20/20	20/20	10/10	10/10	Good	Good	L	R	Excitable.
R. S.	6A	11.8	125	142	132	3.10	2.3	2.8	4.3	5.8	Excellent	3.2	20/20	20/20	10/10	10/10	Good	Good	R	R	Very social.
L. E.	6A	11.8	124	148	140	4.6	3.10	−.2	3.0	4.4	Fair	2.5	20/20	20/20	10/10	10/10	Fair	Poor	L	R	Unsocial; solitary; sensitive. Former teachers had no idea child was bright. Never knew he was in the room.
M. W.	6A	11.7	123	145	130	3.3	3.7	0.9	3.9	5.9	Excellent	2.11	10/20	10/20	9/10	5/10	Good	Good	R	R	Sensitive.
K. S.	4A	9.11	120	112	...	1.11	1.9	1.1	1.5	−.6	Poor	1.6	10/20	10/20	9/10	3/10	Good	Good	R?	R?	Sensitive; nervous; well liked. Very anxious to learn. Good concentration. Father principal of high school.

[1] This covers reading, spelling, physiology, hygiene, language usage, literature, history, civics, geography, and arithmetic on form W New Stanford Achievement Test. The total averages fall much lower than on the subjects recorded on the chart since physiology is not taught in this school, and fourth grade children have not had formal work in history and only a small amount of geography

BOOK REVIEWS

LORD EUSTACE PERCY, Editor in Chief. *The Year Book of Education 1933.* London: Evans Brothers, 1933. Pp. c–860.

This excellently turned out volume of impressive size contains materials of three unequal orders of value and interest to students of Comparative Education. The most interesting and valuable information is found in the sections of the *Year Book* dealing with education in the British Isles, in particular with elementary education and vocational education in England and Wales. Next in significance are the sections devoted to the British Empire, and the least important are those apportioned—too economically to be of practical value—to foreign lands. If a digression be permitted, one would be tempted to show how the selection of topics and the organization—or to be more exact, lack of it—of material relative to foreign countries makes one appreciate ever more the internationally recognized *Educational Yearbooks* edited by Dr. I. L. Kandel, the distinguished Professor of Comparative Education at Teachers College, Columbia University.

The Reader of the *Year Book of Education 1933* cannot help wondering why, for instance, in the essays on foreign systems of education, Germany is omitted though Hungary is given some space. Or why, in the table of "Fundamental Statistics for Germany" (p. lxxx) not even a faint echo is found of the *Grundschule* (the public primary school obligatory for all the children of Germany between the ages of six and ten, in the case of exceptionally gifted children, between six and nine) which is one of the most striking developments of the post-War educational history of Germany. Or why the few pages allotted to France (pp. 724–732) should be occupied with a brief—altogether too brief to be really instructive—*résumé* of the guiding principles on methods and courses of study of the secondary schools contained in the rather voluminous Ministerial Regulations of September 1925 which in truth can do justice to both the Ministry and to readers only when studied *in extenso.* Or why there has not been included in the table of "Fundamental Statistics for France" (p. lxxxviii) any figure, concerning the selection of students for the public schools above the elementary school level which is a far reaching French experiment in a new application of democracy to education? It may be observed in passing that the French subtitle—*Ecoles Maternelles et Ecoles Primaires Supérieures*—to the title "Elementary Education up to the Age of Thirteen" found in the same table should

have included also the term *Ecoles Primaires Elémentaires* without which the subtitle has no sense. There are one or two more inaccuracies in the use of French terminology in the table. The summary of educational policies of the Bolshevist Government of Russia which is well done on the whole ought to include some mentioning of the state organized atheistic training (*bezbogzniki*) given in the schools to the children and adolescents who under the Soviet laws may not receive any religious-moral instruction even in private.

Though the list of detailed criticisms especially in relation to the foreign lands section of the *Year Book* might be prolonged, it remains true that the defects of the volume as a whole are amply compensated by the good qualities and students of Comparative Education will undoubtedly await with great interest the publication of Yea[illegible]k of Education 1934.

MICHAEL DEMIASHK[illegible]
George Peabody College for Teachers.

LEONARD V. KOOS and GRAYSON N. KEFAUVER. *Guidance in Secondary Schools.* New York: The Macmillan Company, 1932. Pp. Xl + 640.

Guidance in Secondary Schools is an exceedingly thorough text which, in my opinion, will be the standard for comparison for some years to come. From beginning to end it impresses the reader with the inclusiveness of its plan, the extent of detailed researches recorded, and the conviction impelled by the sure but careful manner in which the book is written. It is all that the authors state as their intention in the preface—a systematic and comprehensive treatment of guidance in secondary schools (junior high school through junior college).

In the first chapter the authors, beside indicating the need for and describing the scope of guidance, define the term in a manner which is needed. Their definition clarifies the field. It removes the source of much confusion and ineffective effort. A service is done in distinguishing guidance from the whole of education. Two main phases of guidance are distinguished (1) the distributive—the aim is to distribute youth as effectively as possible to educational and vocationla opportunities, and (2) the adjustive—to help the individual to make the optimal adjustment to educational and vocational situations. There is an angle to the second part of the definition which is implied only by the broadest interpretation but which is not omitted by the

authors in their discussion. Mention is made of it here for the reason that there is every indication it will assume a more important place in the near future in the concept of guidance. I refer to personality adjustment. The concept is treated by the authors who append good references in their bibliography. Mental hygiene is in its childhood. As it grows, we may expect to see it assume a greater importance in the adjustive phase of guidance.

The treatise is divided into four main parts "respectively concerned with (1) informing the student concerning educational and vocational opportunities, (2) securing information concerning the student, (3) guiding the individual student, and (4) organizing guidance service." The core of the content of the book was supplied from data given by more than four hundred secondary schools in nearly all of the states. Four year high schools, junior high schools, senior high schools and six-year schools were included. The range of enrollments was from 47 to 4072 pupils. School administrators and guidance directors will find the fourth section most helpful. Six chapters are devoted to The Problem of Organizing Guidance Service, Deans of Girls and Boys, Home-Room Advisers, Advisers and Counsellors, Other Guidance Funcionairies and Relationship, and Organizing Guidance Service in the Junior College. But the discussion does not end there. Educators in the field would find the splendid, comprehensive discussion throughout the book somewhat overwhelming without the final chapter on the Program of Guidance. Items for a program of guidance within a school are given in the order in which they are treated in the book. It is stated that some proposals are not sufficiently evaluated but "the proposals are made in the light of what appears to be the best judgment and best evidence available." The authors are to be commended for making these concrete proposals which emerge from their discussion. It is a characteristic which greatly increases the value of the book to principals, guidance directors and other school administrators.

The text is adapted to classroom use for students of guidance. Summaries occur in many chapters. Questions based on the discussion are given at the end of the chapter and are followed in each case by excellent bibliographies. The publishers have given the book a dignified appearance and have succeeded in combining a good binding with a book of moderate size, one which does not appear to contain the nearly six hundred fifty pages in it.

J. H. Coleman.
Huntington, N. Y.

F. M. EARLE. *Methods of Choosing a Career.* George G. Harrap and Co., Ltd., London: 1931, Pp. 334.

This book is a combination of a research report and a manual on methods of vocational guidance. The frame work of it is an experiment in which six hundred London children were given a guidance program based on psychological tests and related scientific devices while six hundred other students were used as a control group. The results and outcomes of the study are important from a research point of view, while the detailed descriptions of the guidance program, and the methods of carrying it on, furnish much that should prove helpful in other schools.

The principal measures of the value of guidance were in the continuity of work on the jobs taken after leaving school, and the employers' reports. These results favored the group which was given the guidance. The lack of *striking* results, however, suggests that the guidance problem is still far from solved, and that we shall have to give it a great deal more study than we have so far if we are to get every boy and girl properly installed in his niche in life.

C. C. CRAWFORD.

University of Southern California.

EDWARD J. v. K. MENGE. *Jobs for the College Graduate in Science.* Milwaukee: The Bruce Publishing Co., 1932, Pp. VIII + 175.

Vocational and educational guidance have become recognized fields of service, and have taken a variety of different forms. Courses in occupational orientation have been given in junior and senior high schools, and many textbooks on the choice of an occupation have been published. The author of this volume, however, has rendered a unique service in a very specialized portion of the total field by going very thoroughly into scientific occupations for college graduates.

Perhaps the aimlessness of much college work is due to the lack of any realization, on the students' part, of where the whole thing leads. Why major in science? Or in anything else? What chance is there for capitalizing on the intensive study at the end of it all? College students ask these questions about all the departments, and Menge has answered very fully with respect to science.

Perhaps the picture is painted a bit too brightly in spots. Perhaps the book does have a bit of propaganda in it. One gets the idea that

the author believes wholeheartedly in the value and importance of science, and that he would like to add a few converts by displaying in the most attractive form that which his beloved field has to offer. But enthusiasm is praiseworthy; and it is doubly so when there is as much solid substance, substantial fact, and genuine helpfulness back of it as one finds in this book.

C. C. CRAWFORD.
University of Southern California.

REAGAN, G. W. *Fundamentals of Teaching.* Chicago: Scott, Foresman and Co., 1932, Pp. VI + 554.

For genuine teachability this volume will rank very high in comparison with books for education courses. The writer has been very keenly conscious of the needs of *students* throughout, instead of thinking too much about the critical reactions of other education professors. The work is essentially a textbook for a three-hour course in general methods, but it is something more. It is also a set of lesson plans for the professor who teaches the course.

The usual difficulty with methods books is the fact that the students may memorize the chapters or may *discuss* points therein, but that there is nothing very concrete, tangible, and definite for them to *do* as a means of learning the content of the course. The Reagan book not only has general exercises and references at the ends of the chapters, as most books do, but it also stops from time to time in the midst of each chapter to introduce an exercise, project, problem, or experiment to form the basis for application of principles just stated. These suggested applications are quite ingeniously constructed, and appear to be the outgrowth of actual class try-outs rather than mere armchair imagining.

The selection of content for the book is reasonably comparable to that of the better books on general methods. The four parts are on the steps in the teaching process, general pedagogical principles and problems, types of teaching procedures, and extra-instructional aspects of teaching.

The book impresses the reviewer very favorably.

C. C. CRAWFORD.
University of Southern California.

p. 401 after p. 320